The IUPHAR Compendium of Basic Principles
For Pharmacological Research in Humans

Patrick du Souich, Michael Orme and Sergio Erill, eds.

Contributors

Bo Abrahamsen, MD, Ph.D.
Santiago Arroyo, M.D., Ph.D.
Ettore Beghi, M.D.
Yoav Ben-Shlomo, MBBS, Ph.D.
Donald Birkett, Ph.D.
Kim Brixen, M.D., Ph.D.
Kim Brøsen, M.D., Ph.D.
Rosario Calvo, Ph.D.
David W. Chadwick, M.D.
Piu Chan, M.D., Ph.D.
Lara Chayab
Imti Choonara, M.D.
Michel Clanet, M.D.
Giancarlo Comi, M.D.
Per Damkier, M.D., Ph.D.
Simon de Denus, B. Pharm, M.Sc.
Patrick du Souich, M.D., Ph.D.
Anique Ducharme, M.D., M.Sc.
Robert Dufour, M.D.
Kim Fisher, Ph.D.
Jacqueline French, M.D.
Christopher G. Goetz, M.D.
Marie-Claude Guertin, Ph.D.
Neil Hagen, M.D.
Boulos Harauyi, M.D.
Anh Ho Ngoc, Ph.D.
Moustapha Kassem, M.D., Ph.D.
Amos Korczyn, M.D., M.Sc.
Teresa Kus, M.D., Ph.D.
Maxime Lamarre-Cliche, M.D.
Pierre Larochelle, M.D., Ph.D.
Nerea Leal, Ph.D.
Jacques LeLorier, M.D., Ph.D.
Giancarlo Logroscino, M.D., Ph.D.
Fred Lublin, M.D.
John C. Lukas, M.S., Ph.D.
Johanne Martel-Pelletier, Ph.D.
Vittorio Martinelli, M.D.
Richard Mattson, M.D.
Douglas C. McCrory, M.D.
Claudio A. Naranjo, M.D.
John Oldenhof, Ph.D.
Wendy Parulekar, M.D.
Jean-Pierre Pelletier, M.D.
Emilio Perucca, M.D., Ph.D.
Chris H. Polman, M.D.
Jean-Pierre Raynauld, M.D.
Monica Rodriguez, Ph.D.
Myroslava Romach, M.D.
Edward M. Sellers, M.D., Ph.D.
Folke Sjöqvist, M.D., Ph.D.
Per Solberg Sorensen, M.D.
T.J. Steiner, Ph.D.
Elena Suarez, Ph.D.
Caroline M. Tanner M.D., Ph.D.
Jean-Claude Tardif, M.D.
Peer Tfelt-Hansen, M.D., Ph.D.
Jake J. Thiessen, Ph.D.
Bozidar Vrhovac, M.D.
Tom Walley, M.D.

Published by
IUPHAR
Administrative Office
Department of Pharmacology
College of Medicine
University of California, Irvine
Irvine, CA 92697-4625
www.iuphar.org

Cover design by Lindsay Hart

© International Union of Basic and Clinical Pharmacology, 2004
All rights reserved. No part of this publication may be reproduced, stored in a retrieval system, or transmitted in any form or by any means, electronic, mechanical, photocopying, recording or otherwise, without the prior permission of the publishers.

ISBN 0-9533510-6-X
Patrick du Souich
Professor and chairman of the Department of Pharmacology of the Medical School at the University of Montréal, Québec. Vice-chairman of the Clinical Division of IUPHAR.

Sergio Erill
Director of the Esteve Foundation, a nonprofit institution aimed at fostering communication among scientists. He has been Professor and Chairman of the Department of Pharmacology at two different Spanish universities, and Visiting Professor at the Université de Montreal. He has been president of the Spanish Society for Pharmacology and served in the Selection Committee for Merck Sharp and Dohme International Fellowships in Clinical Pharmacology.

Michael Orme
Emeritus Professor of the University of Liverpool, United Kingdom. He was Professor of Pharmacology and Therapeutics there from 1984 to 2001 during which time he was secretary of the Clinical Section of IUPHAR from 1987 - 1993. He was dean of the faculty of medicine at Liverpool University from 1991 - 1996. He was honorary secretary of the European Association for Clinical Pharmacology and Therapeutics (EACPT) from 1993 - 2003 and is currently chairman of EACPT.
The IUPHAR Compendium of Basic Principles for Pharmacological Research in Humans

Introduction ...............................................................................................................................................4

Chapter 1. Ethical Considerations ...............................................................................................................5
Chapter 2. Good Clinical Practice ..............................................................................................................11
Chapter 3. Assessment of Endpoints: Kinetics and/or Dynamics .............................................................16
Chapter 4. Pharmacogenetics and Pharmacogenomics .............................................................................27
Chapter 5. Paediatric Drug Research .........................................................................................................34
Chapter 6. Phase I (Human Pharmacology) ................................................................................................40
Chapter 7. Follow-Up of Drugs After Market Entry ..................................................................................50
Chapter 8. Bioavailability and Bioequivalence ...........................................................................................55
Chapter 10. Drug Utilization .......................................................................................................................76

Pharmacological Research in Cardiovascular Disorders ....................................................................86

Chapter 11. Hypertensive Vascular Disease .............................................................................................87
Chapter 12. Lipid Lowering Agents .........................................................................................................106
Chapter 13. Anti-atherosclerotic Drugs .....................................................................................................118
Chapter 14. Heart Failure ........................................................................................................................132
Chapter 15. Arrhythmias ..........................................................................................................................143

Pharmacological Research in Neurologic Disorders ..........................................................................164

Chapter 16. Epilepsies and Convulsive Disorders ....................................................................................165
Chapter 17. Headache Disorders .............................................................................................................183
Chapter 18. Alzheimer’s Disease and Other Dementias ...........................................................................212
Chapter 19. Parkinson’s Disease and Other Extrapyramidal Disorders ..................................................228
Chapter 20. Multiple Sclerosis and Other Demyelinating Diseases ......................................................248

Pharmacological Research in Mental Disorders ..................................................................................263

Chapter 21. Mood Disorders ...................................................................................................................264
Chapter 22. Anxiety Disorders ...............................................................................................................272
Chapter 23. Schizophrenic Disorders .....................................................................................................278

Pharmacological Research in Joint Disorders ......................................................................................293

Chapter 25. Osteoarthritis/arthrosis Short Term Studies .........................................................................294
Chapter 26. Osteoarthritis/arthrosis Long Term Studies .........................................................................308
Chapter 27. Rheumatoid Arthritis ..........................................................................................................318

Pharmacological Research in Other Disorders ....................................................................................337

Chapter 28. Neoplastic Diseases ..............................................................................................................338
Chapter 29. Analgesic Drugs for Cancer Pain Management .....................................................................352
Chapter 30. Drugs Used in Osteoporosis .................................................................................................366
Introduction

One of the first recorded human trials was conducted by reverend Edward Stone who found in 50 patients that 1.8 g of powder of willow bark cured their fever, results that were published in 1763 (1). The active compound, salicylic acid, was synthesized only in 1860. Since then, innumerable compounds have been used to cure almost any ailment without evidence of activity. Evidence of drug efficacy was initially required in 1962 with the passing of the Food, Drug and Cosmetic Act by the United States Congress. Currently, in all countries, development and approval of new pharmaceutical entities requires controlled trials proving efficacy. In order to standardize drug registration and approval of drugs, the first International Conference on Harmonization (ICH) was held in 1990. Even if a tremendous progress has been achieved by using ICH guidelines, many aspects of human research remain controversial (2), and even for theoretically rather simple trials, such as those aiming at proving bioequivalence, specifications and study methods differ slightly from one to another in different countries (3).

Should we be concerned with refining the methodology of clinical trials? The answer is yes. Let us consider digitalis. William Withering transformed digitalis from a folk remedy to a modern drug when he transformed a "family receipt for dropsy" that contained more than 20 substances, to a single substance by assuming that foxglove was the active ingredient. Clinical observations enabled Withering to recognize the plant’s narrow margin of safety and the importance of dose: just enough foxglove to cause diuresis, but not enough to cause vomiting or very slow pulse. With these observations, Withering introduced foxglove to the medical profession in 1785 (4). Despite many small trials, it took two centuries to clearly demonstrate the benefits of digoxin in heart failure, and we know now that these benefits include reduction of symptoms, improvement in NYHA class, increased exercise time, modest increased in left ventricular ejection force, enhanced cardiac output, and decreased hospitalizations, and that digoxin does not reduce overall mortality but reduces the rate of hospitalization (5,6).

How to conduct trials to demonstrate drug efficacy? Despite the fact that guidelines for drug development are rather standardized, there is less information about the design of a clinical trial. The objective of this Compendium is to provide the scientific community interested in human research with an easy-to-use reference on how to design a research protocol to assess the effectiveness of a drug in a series of pathological conditions.

The Compendium cannot cover every class of drug and condition, and thus it has primarily focused on cardiovascular and nervous system drugs. The section dealing specifically with the design of clinical trials, chapters 11 to 30, is presented according to a common template to facilitate its consultation. This section is preceded by shorter chapters dealing with general concepts that are applied to the development of almost any drug. The Compendium does not intend to constitute a guideline, but rather an easy source of information on how to design and conduct a clinical trial aiming at demonstrate drug efficacy.

The Editors

Chapter 1. Ethical Considerations

Bozidar Vrhovac, M.D.

Professor Emeritus
Medical School Zagreb
Zagreb
CROATIA

and

Co-chairman
National Bioethics Committee for Medicine of Croatia

and

Corresponding Member
Croatian Academy of Arts and Sciences
I. INTRODUCTORY REMARKS

One of many characteristics of modern society is a pronounced interest in ethical questions. Medicine, especially research on humans, is expectedly, at the top of the list. Why is that so? In spite of the fact that many patients receive therapeutic benefits from participating in clinical trials, benefits that may even be greater than those of standard medical care, randomized clinical trials differ from standard medical treatment in their purpose, characteristics, justification of risks and allocations of interventions according to chance. The research based on various interventions potentially poses risks to the participants that are not always compensated for by medical benefits but that are justified by the potential scientific value of the knowledge which will be got from the trial.

The history of international instruments on ethics is not long. Already before World War II, use of controlled clinical trials was proposed and accepted as the scientific, reliable way of proving efficacy and safety of new therapeutic agents. The atrocious experiments performed by nazi physicians during World War II led, almost immediately (1947) after the war, to the preparation of the Nuremberg Code on ethics of medical research. The Helsinki Declaration followed in 1964 and is now (sixth revision in 2000) taken as the gold standard for research ethics, intending to provide a universal set of principles, which direct the ethical conduct of clinical medical research involving human subjects throughout the world. This is still true in spite of several weaknesses which are at the moment of writing these lines intensively discussed. Other instruments must be mentioned, such as the UN General Assembly Universal Declaration of Human Rights in 1948 and the International Covenant on Civil and Political Rights in 1966. The Belmont Report, elaborated in the US in 1979, is in this country very important for developing new drugs. The Belmont report is, in the US, almost better known than the Helsinki Declaration, and, with its legislative revisions performed later, it is still a very comprehensive instrument.

Recent and very important instruments are the documents issued by the 1990 founded International Conferences of Harmonization (ICH) founded in 1990. Originally the basic aim of ICH was to harmonize the requirements for new drugs in the three biggest drug developers, namely US, European Union and Japan. Later, its guidelines and consensus documents have been accepted by the «rest of the world» most probably because of the importance of the pharmaceutical markets in these countries rather than because of other less materialistic reasons.

The first principle of ICH (taken from WHO GCP in 1995) states: "Clinical trials should be conducted in accordance with the ethical principles that have the origin in the Declaration of Helsinki, and that are consistent with Good Clinical Practice (GCP) and the applicable regulatory requirement(s)".

The booklet prepared by the Council of International Organizations of Medical Sciences (CIOMS) in 1982, in 1993 and in 2002 is today the most informative and comprehensive source of information for ethical research, trying to correct the inconsistencies of the latest revisions of the Helsinki Declaration. These inconsistencies perhaps explain why «studies that breach the provisions of the Helsinki Declaration are still commonly conducted, with the full knowledge of regulatory agencies and institutional review boards».

Of the realistic and justified aspects of the Helsinki Declaration, the most important is the respect for the person’s rights e.g. personhood of subjects, followed by investigators beneficence for subjects participating in the trial and distributive justice in distribution of risk and benefit associated with medical research. The growing importance of persons rights (patients and healthy volunteers) is illustrated by the special attention given to trials in vulnerable and socially unprivileged patients.
Issues of conflict of interest, of transparency and of publishing negative trials are closely linked with ethics as well.

II. THE ETHICS COMMITTEE

The guarantee for the ethical conduct of the study should be a multidisciplinary ethical body called in various countries the Ethics committee, institutional review board, independent ethical committee (EC/IRB). Its size is according to many documents of “at least five” members. This important detail is not mentioned in the Helsinki Declaration. With a small number of members it cannot be expected that an institutional review board will be independent when it decides about resources brought by the sponsor to the institution and its investigators. The number of members must be large enough to ensure that besides the layman, the nurse, the ethicist, and the statistician (who are often named as useful non-scientific members of EC/IRB), at least some members must be experts in the medical and scientific aspects of the clinical trial. Scientific and ethical review cannot be separated. How can someone discuss the ethics of a clinical trial without knowing in detail all facts about the disease in question and its standard treatment? Only medically and scientifically competent members of the ethical committee can safeguard the rights, safety and well-being of the research subjects.

Is it optimal that the same ethics committee, and this is often the case, evaluates research projects and other relevant ethical questions which are constantly present in a health institution, such as artificial prolongation of life of irreversibly sick patients, abortions, unethical behaviour of medical staff, to name only a few.

The EC/IRB should be an independent body, either regional or (for smaller countries) central-national and should discuss research projects only. Such development goes in the described direction and many (even bigger) countries already have central ethics research committees or institutions (US Office for Human Research Protection, U.K. Central Office for Research Ethics Committees, Canada National Council on Ethics in Human Research). In the EU for multicentre trials one member country must give one opinion. This is achieved in various ways one of them being that the central committee delegates the decision to a regional one.

II.1. Informed Consent Document (ICD)

How does the EC/IRB functions and what are the foci of their activity? The most important ethical aspect of the clinical trial is the Informed Consent Document (ICD). The already mentioned International Covenant accepted by the United Nations Assembly in 1966 stresses that « no one shall be subjected without his free consent to medical or scientific experimentation ». A number of documents, meetings and discussions have been written and organized about the optimal format of this important ethical aspect of the clinical trial documents.

It is of the utmost importance that the subject participating in a scientific research project, for instance a clinical trial, understands all details of the planned experiment. To reach this aim the investigator must ensure that the prospective subject has got all the necessary information on the basis of which he reaches at the decision to take part in the trial without having been subjected to coercion, undue influence or inducement, or intimidation. The ICD should contain a statement indicating that the study involves research, should describe the purpose of the research, the expected duration of the subject’s participation, should contain the description of the procedures to be followed (with the indication of which are experimental), of the study treatments, and, if applicable, the nature of random assignment. Moreover the ICD should contain a description of the foreseeable risks or discomforts, of the benefits that may be expected for the subject or for others and disclosure of appropriate alternative procedures or courses of treatment, if any, that might be advantageous to the subject. In addition the ICD will include a statement
describing the extent to which confidentiality and privacy (of records identifying the subject) will be maintained, a statement that participation is voluntary and that the refusal to participate involves no penalty or loss of benefits. It must describe the actions foreseen in the case of injuries (compensation, medical treatment), explanation on whom to contact for additional questions, and, if applicable, any other necessary detail which ensures complete comprehension of above mentioned points by the participant.

Whenever it is not possible to obtain the written ICD, the non-written consent should be documented and witnessed. The problems of obtaining a written ICD for trials in patients with decisional (cognitive) impairment such as patients who are mentally ill, those with Alzheimer’s disease, acutely ill subjects (head trauma, cardiopulmonary arrest, and stroke) or children (no uniform criteria for assent and dissent exist) are considerable. In these cases a legally authorised representative, a proxy or an advanced ICD (in the case of anaesthesia for example) should be used.

In addition, financial details of the trial should be disclosed. Transparency of financial arrangements encourages people to do the right thing. The individual investigator should not stand to benefit personally in financial terms from their involvement in the study.

**II.2. Analysis of Details**

Another function of the EC/IRB is to conduct a careful analysis of all details, which could influence the reliability of the trial results. The new sentence in the latest revision of the Helsinki declaration must be mentioned here (articles 19 and 20): “Medical research is only justified if there is a reasonable likelihood that populations in which research is carried out stand to benefit from the results of the research”.

The analysis should begin with the investigator and their team, their potential to recruit subjects without aggressive behaviour, the likelihood of a conflict of interest when trying to serve both the best interests of the patients and the best interests of the research. The analysis should also consider the remuneration and other advantages for both the investigator and their institution. The patient selection, the planned measurements (frequency, justified invasiveness), concomitant and rescue therapy, monitoring of adverse events (especially those which will indicate the need to stop the trial) and comparator therapy must be analyzed.

**II.3. Placebo**

The function of the EC/IRB is to analyze the need to use placebo. The use of placebo has been considered by many as non-realistic and unjustified. The most controversial item of the last two revisions of the Helsinki declaration (article 29) states: “The benefits, risks, burdens and effectiveness of a new method should be tested against those of the best current prophylactic, diagnostic and therapeutic methods. This does not exclude the use of placebo, or no treatment, in studies where no proven prophylactic, diagnostic or therapeutic method exists”. Strictly interpreted, this article would rule out the use of placebo controlled clinical trials e.g. a dummy treatment administered to the control group in a controlled clinical trial in order that the specific and non-specific effects of the experimental treatment be distinguished whenever licensed therapeutic method already exists. Active controls (which the investigators are keen for) cannot, in many circumstances, provide reliable evidence of efficacy and safety of the new drug, except by showing the non-inferiority of the new drug. There are many groups of therapeutic agents where placebo controls are justified or even mandatory: analgesics, many psychopharmacologicals, antihypertensives, antianginals, antiarrythmics and drugs used in primary prevention to name only a few. It is essential that the use of placebo does not pose a risk of serious discomfort, irreversible harm or death or that existing therapy improves survival or decreases serious morbidity.
The sixth revised version of the Helsinki Declaration raised a number of discussions, many more than after the fifth version in which the same proposal was already present, with the result that the World Medical Association prepared a special footnote which states: “The WMA hereby reaffirms its position that extreme care must be taken in making use of placebo-controlled trials and that in general this methodology should only be used in the absence of existing proven therapy. However placebo-controlled trials may be ethically acceptable, even when proven therapy is available, under the following circumstances:
- where for compelling and scientifically sound methodological reasons its use is necessary to determine the efficacy and safety of prophylactic, diagnostic or therapeutic method; or
- where a prophylactic, diagnostic or therapeutic method is being investigated for a minor condition and the patients who receive placebo will not be subject to any additional risk of serious irreversible harm”.

Beside WMA, many meetings in various international organizations like the European Medicines Evaluation Agency (EMEA), the Food and Drug Administration (FDA), the Pharmaceutical Research and Manufacturers of America (PhRMA), the International Federation of Pharmaceutical Manufacturers Associations (IFPMA) and the European Forum for Good Clinical Practice (EFGCP) have discussed the need for a new revision of the Helsinki declaration because of the presence of other conflicting articles. For example, article 30 states that “every patient should be assured of access to the best proven method identified by the study”. This is not acceptable because of the widely known fact that one study cannot identify “the best proven method”.

Therefore, a new version of the Declaration of Helsinki is to be expected. In September 2003 the General Assembly of WMA founded a Working group (its deadline is May 2004) with this goal.

II.4. Payment

Payment in the form of money, gifts, and privileges can only be offered as a recruitment incentive not as a benefit for participation. In clinical trials the prospects of benefit from an experimental treatment and the provision of free ancillary care are viewed as compensation for participation. Healthy volunteers do not need treatment and care. So payment is justified as an incentive for participation.

There are many other ethically sensible areas of clinical research. The examples are, beside those already mentioned, the need to withhold treatment, wash out periods, research involving foetuses and in vitro fertilization, involving pregnant women, children, college students and prisoners. In each of these cases the local EC/IRB has to adapt its decisions according to the local legislation, uses and environment.

In conclusion, clinical trials have numerous ethical aspects. A scientifically and medically well planed clinical trial is ethical and represents the only way for obtaining reliable results which will help in better treatment of a wide circle of patients. Local EC/IRBs have to define what is methodologically essential and ethically appropriate and these aspects are still the subject of intense debate. Ethical committees structured as proposed in this chapter guarantee that ethical principles, accepted to day as appropriate, are observed.

III. SUGGESTED READINGS

11. The European Agency for Evaluation of Medicinal Products: EMEA/CPMP, Position Statement on the USE of Placebo in Clinical Trials with regard to the revised Declaration of Helsinki, EMEA/17424/01.
Chapter 2. Good Clinical Practice

Per Damkier, M.D., Ph.D.
Clinical Pharmacology
Institute of Public Health
University of Southern Denmark
Odense University
Winslewparken 19
5000 Odense C
DENMARK
I. INTRODUCTORY REMARKS

GCP, Good Clinical Practice, is an international set of ethical and quality standards that applies to medicinal trials in humans. Compliance to these standards provides the health authorities as well as the general population with assurance of the integrity of trial subjects and the validity of the data generated. It is important to stress that these standards de facto relate to research; in fact a more appropriate title would be Good Clinical Research Practice. These standards have been in effect for many years pertaining to all trials intended to generate data for marketing authorisation procedures whether it be new medicinal products or a line extension for an already marketed product. As a general rule, academic research involving already marketed products and not intended to generate results for marketing authorization purposes has been exempt from these rules. As more general attention and political focus is given to quality assurance and to the autonomy of the individual within health care systems, these standards are getting more attention world-wide; Within the EU, these rules, as of May 1st 2004, pertain to any medicinal trials in humans, including those involving already marketed drugs and trials performed without industry engagement. This represents a serious challenge to the academic independent drug related research, as systems to assure GCP compliance must be developed, which in turn requires allocation of appropriate resources.

I.1 History

The first documents describing some quality recommendations for the design of clinical trials in humans can be dated back to USA, where Harry Gold from Cornell University Medical School published two influential papers Conference on Therapy in 1946 and again in 1954. The emergence of the thalidomide disaster around 1960 further served to justify the need for formal guidelines and procedures for clinical trials with new medicinal products. Throughout the 60’s and 70’s guidelines were refined and implemented throughout the world. These were, unfortunately, far from being easily comparable due to differences in approaches to clinical trials, mainly between USA, Japan and Europe. It became obvious that some sort of international consensus on this issue was overdue in order to promote mutual recognition of clinical trials and marketing authorization procedures. The result was the birth of the current guidelines that in effect are known as ICH (International Conference on Harmonization)-GCP.

I.2. GCP concepts

In order to give a better understanding of the principal GCP concepts, some key definitions are given below:

Sponsor
A person, institution, company or organization which takes responsibility for the initiation, management and/or financing of a clinical trial. Note that the sponsor does not necessarily finance the study. Sponsor and investigator may be identical in which case the term (surprise) sponsor-investigator is used.

Investigator
A person who is responsible for the trial conduct at the trial site. If multiple investigators are involved at a trial site, a principal investigator must me appointed.

Monitoring
The act of overseeing a clinical trial to ensure that it is conducted recorded and analysed according to the trial study plan, standard operating procedures, GCP and regulatory requirements.

SOP, Standard Operating Procedure
A set of written detailed instructions to achieve uniformity of the performance of a given function.
Audit
An independent examination of all trial related activities and documents in order to determine if the trial was conducted, recorded and analysed according to the trial study plan, standard operating procedures GCP and regulatory requirements. The audit procedure is independent of the monitoring procedure.

II. GCP IN CLINICAL RESEARCH

The ICH-GCP guideline specifies, in rather general terms, how to design, conduct, record and report a clinical trial in accordance with GCP standards. In order to comply with this, it is the responsibility of the sponsor, or sponsor-investigator, to ensure that a set of standard operating procedures is written. An important point is that the guideline explicitly states that the level and intensity of the GCP monitoring process is specified by the sponsor according to the complexity of the study: size, purpose, design, blinding, outcome measures etc. These standard operating procedures form the basis of the central element of GCP-compliance: the monitoring process. The monitoring process must be conducted by individuals possessing documented skills of GCP monitoring, and, obviously, the monitors cannot be directly involved in the study otherwise.

II.1. The monitoring process

The monitoring process is the fundamental aspect of GCP compliance. This systematic process is based on study-specific SOP’s, and serves to document that the study complies with GCP standards. There are three phases of this process some core issues of which are listed below. Please note that this is not an exhaustive listing.

**Before** study initiation the monitor visits each investigation site to verify that
- Sponsor and investigator responsibilities are properly described
- Relevant authorizations are present
- Written informed consent is acceptable
- The investigation site realistically can provide the specified number of subjects within the specified time-frame
- Procedures for handling of trial medicine and laboratory tests are present
- Source documents are specified

**During** the trial, the monitor visits the trial sites to verify that
- The trial is performed and documented as planned
- Informed consent is given from every participant
- In- and exclusion criteria are fulfilled
- Data are correctly recorded and are in accordance with the specified source documents
- Corrections in the case report form (CRF) are properly performed and documented
- Serious adverse event are handled correctly

**After** the trial is completed the monitor verifies that
- Trial medicine is accounted for in detail
- The trial database is properly secured and validated
- Trial documents are filed and stored properly

In effects this means that quite a number of SOP’s must be developed and maintained. A standard operating procedure must exist for all items and procedures below:
- Protocol
- Informed consent
- Investigators brochure
All monitors’ visits must be accompanied by a written report describing what was monitored, the outcome of the monitoring process, including errors and deviations, and initiatives to correct the latter. For the individual researcher this likely represents an insurmountable task and flexible systems to handle GCP in independent academic drug research must be developed.

II.2. A Danish example

In Denmark a public GCP monitoring service has been organised, and a brief overview is given here for inspiration: The first public GCP initiative was taken back in 1995 at the University of Aarhus. Anticipating the implementation of the aforementioned EU directive, it was estimated that about 80 independent academic drug trials in humans were initiated every year in Denmark. A coordinated activity and initiative has resulted in the presence of three GCP-units in Denmark, which are all situated around medical schools and universities. The GCP-units are partly funded by the Government; we are estimating that, having reached steady-state, about 20 full time monitors will be employed, and that a total budget would linger around USD 2 million. The three units work closely together and have agreed upon identical SOP systems, and apply identical principles of services: Smaller trials, typically trials in Ph.D. projects, are monitored free of charge, while larger trials must account for factual GCP-costs, if more than 100 hours of service is required. Despite initial skepticism from researchers this system has so far proven manageable, but the full scale test still awaits, at the time of writing, the implementation of the EU directive.

II.3. Challenges and perspectives for independent academic drug research

The challenge is to assure compliance with GCP standards, while not consuming an insurmountable amount of resources, which would cripple the independent academic clinical drug research. It is here that one must recall that the guidelines are general specifications of GCP standards. However, the translation into factual procedures, and the specific way these are implemented leave some room for breathing and interpretation. The approach that the pharmaceutical industry and most CRO organisations have taken is meticulously detailed and very resource consuming. This is understandable, as they cannot afford to have a large pivotal phase III trial be subjected to scrutiny, or even rejection, by health authorities due to a GCP related issue. Hence, the industry developed GCP concept is designed to account for a worst case scenario. So as GCP seems to enter an era were the principle is likely to become applicable to any drug related research in humans, it is really up to all of us, as independent academic clinical researchers, to influence the way that GCP is implemented. There is no denying that implementation of GCP in academic research programs will assimilate some of our scarce funding. However, it is the opinion of this author, that we should welcome many of the principles covered by GCP. Quality in performance and respect for patient’s integrity is imperative for the thrust on which the very existence of any health care system relies, and there are simply no valid arguments to exclude research related activities from these principles. And, if we make our opinion and experience heard it is my belief that it is possible to reach an interpretation of the
GCP guidelines which satisfies this principle and still allows for a fruitful continuation of independent drug research.

**III. REFERENCES**


The FDA has an entire homepage with plenty of easily accessible information: [http://www.fda.gov/oc/gcp/default.htm](http://www.fda.gov/oc/gcp/default.htm).
Chapter 3. Assessment of Endpoints: Kinetics and/or Dynamics

Rosario Calvo, Ph.D.¹

John C Lukas, M.S., Ph.D.²

Elena Suarez, Ph.D.³

Monica Rodriguez, Ph.D.³

Nerea Leal, Ph.D.³

¹ Department of Pharmacology
University of the Basque Country
Leioa, Vizcaya 48940
SPAIN
Tel    + 34 94 6012761
Fax    + 34 94 4800128
Email   rcalvo@lg.ehu.es

² Department of Pharmacy and Pharmaceutical Technology
University of Salamanca
Salamanca 37001
SPAIN

³ Department of Pharmacology
University of the Basque Country
Leioa, Vizcaya 48940
SPAIN

Corresponding Author:
Rosario Calvo, Ph.D.
I. INTRODUCTORY REMARKS

A crucial, yet often limiting, factor in the advance of the majority of sciences is the ability to measure and analyze those variables that are truly relevant to a particular field, as for example, in pharmaceutical science, the concentrations in plasma and the effect of a drug. These measurements would be of little use were it not for the development and progress of collateral disciplines such as pharmacokinetics (PK) or pharmacodynamics (PD). The PK examines the relation between the dose administered and the achieved concentration in blood or at the biophase, or “what the body does to the drug”, and the PD deals with the relation between drug amount and effect, or “what the drug does to the body”. In its turn, progress in PK or PD is due mostly to the availability of sensitive and specific analytical methods in order to determine the evolution of the levels of drug and its metabolites in biological fluids and tissues as well as to quantify the drug’s therapeutic effect.

In parallel, the mathematical techniques and methods used in pharmaceutical science to characterize the kinetic processes have also developed considerably, mostly borrowing from other disciplines such as chemistry, physiology or enzymology. The application of these analytical and mathematical techniques to specific studies has enhanced the knowledge of the PK behavior of many drugs, constituting an important advance in adjustment of the dose. But, unpredictability remains a problem since the kinetics usually show large interindividual variation, mainly due to genetic, environmental and pathophysiological factors. For this purpose, state of the art statistical methods generally denominated as “population” PK/PD models have been introduced in the field aiming to evaluate the between (inter) and within (intra) individual differences in the PK/PD in medicated populations of subjects.

The dose is related to the effect through the PK and also through the PD which relates the concentration with the effect. Comprehension of these relations is truly essential for the rational development of therapeutics because the PD determines the target concentration required to produce a specific effect or endpoint, while the PK specifies the appropriate dose regimen to reach that target. This constitutes the basis of what in therapeutics is called “Target Concentration Intervention” (TCI), in contrast to the therapeutic drug monitoring (TDM) methods based solely on a single concentration and Bayes informed PK.

As in the case of PK, methods have also been developed to assess the pharmacological effect in vivo and provide biomathematical model descriptions of the PD. It is already well known that the interindividual differences in PD are numerous and are related to factors such as age, race or disease. Therefore, future efforts must lead to the exploration of covariate models for predicting individual PD parameters. This knowledge together with its PK counterpart could constitute the basis for rational individualized therapy. The combination of PK and PD models permits predicting the temporal evolution of the effect at any dose or regimen.

Also, PK/PD analysis of concentration – time – effect data assists in detecting a series of underlying complications. For example, tolerance development, the formation of active metabolites or the desynchronization between the evolution of the drug concentration and that of the effect. The latter can be due either to drug distributing into an effect compartment or because of an indirect effect mechanism. In all these conditions the plasma concentrations could not be used directly as “targets” (TDM) but integrated PK/PD models would permit estimation of the drug concentration at the effect site as well as the relationship between them.

II. OBJECTIVE

The main objective of therapy is to achieve maximum efficacy avoiding the risk of toxicity. This can be achieved via empirical adjustment of the dose, based either on observations of effects (effect – time
evolution) and/or concentrations (concentration – time), or optimally based on knowledge of the pharmacokinetic (PK) and pharmacodynamic (PD) parameters.

II.1. Effect – time evolution

The measurement of the pharmacological effect and the adjustment of the dose as a function of the effect appears, in principle, as the most sensible and intuitive approach. However, there are several problems. The observed pharmacological effect, as well as the time it takes to achieve it and its duration, are measurements that change with the dose and the mode of administration, therefore they are variables. This implies that they cannot be used to make predictions in other situations, different from those of the observations at hand.

Additionally, from the knowledge of the effect vs time evolution there can be no direct deduction of parameters, thus having to employ integrated PK/PD analyses as will be discussed below. Parameters are considered those characteristics of the drug that do not change with time, dose or administration route and therefore can be used to adjust or predict adequately any therapeutic regimen. For example, the classical dose - effect relationship provides a useful estimate, that of the dose producing 50% of the maximal effect, ED50. Yet, this is not a true parameter since it depends on the time post-dose when the effect was measured.

Dealing with the pharmacological effect is complicated further because of the difficulty to obtain a precise, objective, and continuous measurement of the effect. The use of biomarkers attempts to relieve this problem although we are still far from having the ideal (and validated) biomarker for the majority of drugs.

The use of biomarkers in drug development and clinical practice was revised in the Ninth European Federation of Pharmaceutical Science Conference on Optimizing Drug Development held in Basel 2001 (1). The terminology has been put up to date and the differences between the terms “end point”, “surrogate endpoint” and “biomarker” clarified. Biomarkers are now defined as “physical signs or laboratory measurements that may be detected in association with a pathologic process and that have putative diagnostic and or prognostic utility”. They are, therefore, factors which can be measured objectively and are evaluated as indicators of biological or pathological processes and/or indicators of the response to a therapeutic intervention. In general, biomarkers have a much wider range than surrogate endpoints. We understand as endpoints those variables which can be used to measure how a patient feels after a specific treatment or how a specific body function evolves, “the clinical impact of therapeutic intervention”. The term surrogate endpoint implies that some variables related to the endpoint or the clinical response have been used as biomarkers but are not the final response of the drug (2,3). For example, blood pressure is a biomarker for prevention of hypertension and, at the same time, a surrogate endpoint for the prevention of myocardial infarction and stroke. The measurement of the degree or percentage of prevention would be the endpoint. This example serves also to demonstrate the difficulty in selecting adequate biomarkers since it is still discussed whether the systolic, diastolic or mean blood pressure is the one of interest. The most widely used biomarkers are plasma concentrations of drugs that are used as guides to dosage in clinical practice (e.g. TDM).

Biochemical and molecular biomarkers, such as leukotrienes, angiotensin I and II, or CD4 cell count are of great utility but are hampered by complexity in their mechanism of action which is widely interconnected to other processes. Consequently, no single such biomarker can predict a significant proportion of the observed clinical endpoint. For example, CD4 cell count explains 30% of the survival to HIV and CD4 count plus viral load explains 70%.

Something similar occurs with the gene biomarker products which are also under rapid development. Genes and their function are identified in the genome. Then the proteome is used to identify proteins from
selected genes. Evaluating how mutations cause disease as a result of protein differences, between healthy and diseased subjects, appears to provide candidate gene biomarkers. However, the complexity of the genome or of the pathway from expression to phenotype to macroscopic reality has deflated initial hope.

Adversities aside, the development of adequate biomarkers for a drug apart from better characterization of its PK/PD for analysis or prediction, now appears crucial for the drug development effort. Valid biomarkers help completing the proof-of-concept in the early phases, facilitating decision on continuation with the new drug. Eventually, biomarkers permit reduction of the number of patients in later phases (II or III) and adjustment of the dose in specific populations or in individuals receiving the drug through different routes or dose regimen. Recall however, that even when the ideal biomarker is known, it is necessary to have the PK/PD parameters permitting to make predictions.

**II.2. Concentration – time evolution (PK)**

The observations of plasma or blood concentration are also non generalizable in their pure form. The maximum concentration reached after a specific dose or the peak or trough after repeated dosing, are measurements useful exclusively for adjusting the dose in situations reproducing the one where they were obtained (body – drug whole). Nevertheless, the measurement of the concentration evolution with time is advantageous, compared to the evolution of the effect, since from this kinetics the PK parameters can be derived that are of great use in adjustment of the dose.

Pharmacokinetics has advanced considerably with the use of mathematical models and computer packages which aid in the analysis and processing of the information generated in clinical practice. The PK also permits the simulation of conditions affecting a particular patient. These models and accessory packages are *tools* which assist in obtaining parameters but add no scientific surplus value to the information than that input by the experiment, including the user. Modelling will simply reflect the knowledge of the physicochemical characteristics of the drug, the precision of the analytical techniques for drug assays. With more complex models, the qualifications of the user also become more important.

The basic or primary PK parameters are apparent volume of distribution (V) and clearance (CL). V is defined as the relation between dose and initial concentration (V = Dose/ Co). CL is a relation between elimination rate (or distribution rate, for intercompartmental clearances) and concentration in blood or plasma. With passive phenomena (or first order kinetics), which are the most usual in the body, a larger dose implies a larger concentration in blood or plasma and consequently larger elimination rate, thus V and CL are always constant. They are therefore considered parameters and permit prediction of the dose which would be necessary in a patient to produce a specific effective concentration level.

For example, knowing V we could predict the dose needed to reach a specific target (loading dose = target concentration x V). Additionally, CL is a parameter independent of the complexity of the kinetic model (mono, bi or tricompartmental) and is calculated simply as Dose / AUC (assuming complete absorption of the drug). In steady state, after an infusion or multiple dosing, CL is related to the steady state concentration (Css) or average Css (Css = infusion rate / CL), so this parameter can be used to predict the dose regimen (maintenance dose = target C ss x CL). Both parameters, CL and V, are primary parameters and are directly related to the physiological processes of the organism. They give us an idea of the relative importance of the space where the drug is distributed and of the organs which eliminate it.

Another PK parameter, commonly employed in dose adjustment, is the elimination rate constant (Kel). It is a mixed parameter that depends on CL and V (Kel = CL / V) and is often recast in the form of the half-life parameter as \( t_{1/2} = \ln (2) / \text{Kel} \), now with units of time. This parameter gives an idea of when the steady state is reached and can be used to predict when a C ss monitoring sample can be taken.
A facet which should not be neglected when performing a kinetic study or using the concentrations as markers is to know a priori what needs to be finally estimated from the observations. The parameters obtained depend on what has been measured, for example, metabolites, unbound or total drug, enantiomers or racemic mixture.

Another frequently encountered issue is the necessity or not to measure unbound drug (not bound to plasma proteins). It is important to recall that on some occasions the unbound concentration in plasma should be considered since binding in plasma and tissues (the effect site) may not be equal. Nevertheless, if the binding is linear in the range of therapeutic or toxic concentrations, the free and total concentration are simple ratios one of the other and in these cases it is not necessary to measure the free concentration. In contrast, nonlinear plasma protein binding (free drug concentrations increase disproportionately with increasing total drug concentrations) can create havoc in analysis unless free drug concentrations are measured. Perhaps an advantage of parameter estimation as a function of free concentration is that it supplies information about the (intrinsic) behavior of the drug excluding possible differences in the degree of binding. For example, the V of unbound drug, Vu, corrected for the weight, can be extrapolated from animal to man permitting estimation of the first dose in humans. An approximate value of CL for the unbound drug, Clu, can be obtained from “in vitro” studies with microsomes in the initial stages of development.

In conclusion, V and CL are fundamental parameters for establishing a dosing regimen, but, depending on the physicochemical characteristics of the drug, the models for the distribution can become complicated and the number of parameters may increase, particularly there may now be more than one volume of distribution. In this case, the volume used to adjust the therapeutic dose, should be the one closer to the effect site (and could also be the steady state volume, Vss) (see Propofol example below). It is important to remember that the basic concepts reflected in V and CL are always applicable independently of the complexity of the complete model.

II.3. Programs for data analysis

Depending on the experiment, the intentions may range from obtaining estimates of PK (PD) model parameters in a single subject or the mean for multiple subjects (a population), or up to the rigorous resolution of the inter and intra subject variability in a population as reflected into statistical distributions of the parameters (Bayesian priors). A subsequent, yet very important task, is usually that of relating the parameters, or their inter subject variability with individual specific covariates, so that a priori prediction of the individual PK (PD) characteristics, and hence the dose, can be improved.

Several software packages have been designed and marketed, mainly in the last 20 years and can be distinguished as single subject analysis programs, or population analysis programs, although both can be used, with the exceptions of some occasions, to perform single or multiple subject analyses. The decision about which approach to use depends, in order of significance, on the density of the observations (rich or sparse data), the scope of the study (e.g. obtaining estimates for a single subject only, or Bayesian priors for a population), and ad hoc criteria regarding modelling.

Rich sampling is when there are several drug concentration observations per individual, necessarily more than or equal to the parameters in the model and well distributed in time (e.g. phase I). Sparse sampling is when there are fewer data points per subject than parameters in the PK model to be estimated. Sparse data occur frequently, e.g. in the clinic, due to monitoring restrictions or in phases II, III or IV, due to logistical or ethical considerations. It is also frequent in dose escalation experiments with drug levels below the quantification limit, particularly in small animals, and also in toxicological single point per animal studies.
With rich sampling, individual fitting programs can always be used in single or multiple subject problems, in addition to population specific packages for the latter case (Fig. 1). In sparsely sampled designs, some knowledge about the population at large is always required for estimation of the individual PK model. In single subject sparse PK model estimates, a Bayesian prior from the population has to be introduced at some point in order to inform or support the algorithm on the data gaps. Single subject PK/PD modelling packages are WINNONLIN (Pharsight Corp., Mountain View, CA) and SAAM (Saam Institute Inc., Seattle, WA), this latter permitting the introduction of Bayesian priors from the literature or earlier studies when the data is sparse (e.g. a monitored concentration for each subject).

When many individuals are treated as one within a single subject fit, the analysis performed is known as naive and produces estimates of the mean parameter without any indication of statistical spread. If that information is desired, with rich data and for more than one individual the single subject model fitting run can be simply repeated or iterated for each case. An improvement is a population – like analysis known as standard two stage (STS), because in the first stage all subject specific PK or PD parameters are obtained and in the latter, their centering and dispersion (mean, standard deviation) are simply calculated. With STS, relationships with covariates (e.g. age, weight, sex, and creatinine) can be assessed with standard commercial statistical analysis packages (SPSS, SAS, S-PLUS etc).

Figure 1. Schematic of decision tree for use of population methods in data analysis.

For sparse data and multiple subject experiments, the use of population or population modelling approach is most appropriate. Belonging to the well known statistical problem of mixed effects, population algorithms use the information from the remaining population to complete the model of each subject in an iterative process of enriching a prior parameter distribution at each step. Population methods eventually produce estimates of the complete distribution for the PK model parameters (population mean, standard deviation), useful as Bayesian priors, as well as the distribution of the residual error: measures of the inter- and intra-individual variability respectively. Population analysis programs are NONMEM (nonlinear mixed effect modelling, NONMEM Project Group, UCSF, CA) (4) or NPEM (non parametric expectation maximization, Laboratory of Applied Pharmacokinetics, USC, CA) (5) and have been validated and compared (6,7). Recently, a package with a more user friendly interface has been introduced (WINNONMIX, Pharsight Corp., Mountain View, CA), and although not widely used so far, it is lately gaining ground (8,9).
Population fits are usually followed by a maximum a posteriori (MAP) Bayesian estimation step, where the PK (PD) model parameters are obtained for each individual based on the just obtained population priors.

Generally, population analyses are far more complex than single subject approaches, in terms of expertise required and man hours and computer time invested, mainly because there is no single pharmacostatistical model solving the parameter estimation problem. Neither is there a single best fit criterion and the solution process includes visual inspection of residuals or evaluation of confidence intervals. Thus the analyst becomes part of a loop in successive model improvement steps. “Turn of the crank” modelling is impossible with population methods. A simple PK model which may take minutes in a computer and a single subject algorithm to solve may require days or weeks for its population counterpart, even excluding computer time delays due to the size of the sample. The time and knowledge invested is largely extended when multiple occasions of the same subjects and covariate models for the PK parameters are created within the package. The complete population analysis, beginning with the design of the samples, collection, analysis, covariate modelling and possibly simulation, is a highly demanding task.

The PK parameters depend on demographic and physiopathological covariates such as renal insufficiency, diabetes, age, sex and weight and this finally affects the dose. The causality or pathways of such variation are usually not well known and thus, unpredictability remains a problem for most drugs. Covariate model development is a very important effort subsequent to any characterization of the individual kinetics in a population. Some population packages allow introduction of a theoretically unlimited number of covariates in the population fit, thus permitting immediate reduction of the inter-individual variability (NONMEM, WINNONMIX); others allow only a limited number of covariates to be introduced in the fit, thus covariate modelling is performed externally.

Even drugs in use for years, like methadone, in clinical practice, show variability in the response (10). In recent experimental work with methadone, it was observed that sex (11), protein binding (12) and P-glycoprotein (13) modify the PK/PD of methadone and it has been suggested that these covariates could be implicated in the variability observed in the clinic. Another example is propofol. Many studies report on the influence of various covariates on the kinetic parameters of propofol, but without a consensus as to the importance of each one (14). Weight, age and formulation have been associated with the variability. Additionally, plasma protein binding, mainly to lipoproteins, is modified in thyroid, diabetic or critical patients, which could finally impinge on the kinetic parameters (15-17). Models have been developed between these variables and the unbound fraction of propofol “in vitro”, which could aid in the inclusion of lipoprotein levels in population PK analyses (17).

Immunosuppressant medication also shows worrisome variability in the kinetics which is further complicated by the standard oral administration. Much of the variability described in the parameters appears to be associated with the bioavailability which may also vary with post transplantation time (18). In this situation, in addition to estimating the mean parameters it is important to quantify the variability, which is important in adjusting the dose and the regimen. A variation in CL, for example, would immediately reflect in a change in the required maintenance dose or the dosification rate. Population covariate modelling often deals with inter occasion variation within the same subject, in addition to the inter intra individual variabilities.

These concepts are important to the Target Concentration Interval (TCI). In NONMEM, for example, the overall variability for each patient can be summarized in three parts, the between subject variability (BSV) and the inter occasion variability (IOV) for each PK parameter and the within subject variability (WSV) for the concentrations. Then covariate models can be developed to reduce the BSV and IOV. If the dose is to be adjusted between different subjects the BSV and IOV must be treated for the parameter of interest (e.g. CL), but if the adjustment is within the same subject the WSV has to be considered.
The population approach has been used for years, and increasingly, in all phases of drug development (19) and in postmarketing studies. Some of the most recent studies with NONMEM, which have implied an advance in the adjustment of the dose in the clinic, are listed in Table 1.

Table 1. Some of the latest studies where population methods are employed for dose adjustment.

<table>
<thead>
<tr>
<th>Drug</th>
<th>Reference</th>
</tr>
</thead>
</table>

Population methods are intimately tied to simulation, deterministic when random components are absent or stochastic when they are not. The latter is used in drug development for “in silico” dosing regimen and risk assessment, facilitating the reduction of trial patients in the later phases. Most of the above packages can be used to perform simulations in a population setting (20).

The applicability of the models, the presentation of concise final reports, and the relevance of the study are factors which should be considered in order for population analyses to be useful in clinical practice and in learning about the behavior of the drug in its development phases (21).

In conclusion, dose predictions are usually simple applications of the elementary PK principles discussed above and permit estimating a target concentration after administration of a specific dose or a particular regimen. Nevertheless, the target concentration should not be far removed from knowledge of the concentration – effect relationship because, in fact, that is where the target originates. The selection of a target concentration requires exploration of the PD relation not only for the desired effect but the range from undertherapeutic to toxic effects. In clinical practice, there is debate regarding the utility of two approaches: TCI which is based on the above discussed PK/PD based principles and Therapeutic Drug Monitoring (TDM) which is based on adjusting the dose to maintain a range of drug levels, perhaps loosing from sight the concept of target and its immediate relation to an effect.

II.4. Concentration – effect relationship (PD)

The concentration – response relationships lead to the estimation of PD parameters, useful in drug development as well as in the clinic. The most common PD parameters are the maximum effect that can be reached (E_{max}) and the EC_{50}, or concentration in plasma or blood capable of producing 50% of the maximum effect, known as potency in “in vivo” studies. These parameters are independent of the dose or time and therefore allow prediction of the effect at any concentration independently of how it was generated. These parameters, like with the PK, are obtained from models, although these are typically diagnostic at steady state rather than with explicit time dependence. The classical models are the E_{max} or hyperbolic model and the sigmoid model which ensue from the classical drug – receptor relations. The E_{max} model is expressed as,

$$E = E_{max} \frac{C}{EC_{50} + C}$$
In this form it is seen that once the parameters are obtained the effect at any concentration (C) can be estimated. Additionally, the parameter EC$_{50}$, expressed as free (unbound to protein) concentration of drug, offers valuable information since it usually is of the same order of magnitude as the “in vitro” potency of the drug, IC$_{50}$ obtained from receptor binding studies in early drug development stages. As such, it can be considered as the target for therapy.

In a recent study with lerisetron, a new 5HT$_3$ antagonist in phase III development, the PD was measured using its surrogate effect, the Bezold-Jarisch reflex. The observed value for the EC$_{50}$ “in vivo” was in the range of the affinity of lerisetron in binding studies published earlier, which allowed the assumption that the activity of lerisetron was due to the parent product and not to the presence of possibly active metabolites, as had been suggested for other compounds of the same family (22). This approach has also been used for comparing adenosine and other lipophilic derivatives as well as various benzodiazepines (23,24).

In spite of the importance of knowing the concentration - effect relationship, for many drugs there is little documentation regarding their PD parameters, including those which are commonly monitored such as aminoglycosides, cyclosporin, phenytoin and digoxin. An exception is theophylline for which the E$_{\text{max}}$ value is known (expiratory flow rate) and as well as the EC$_{50}$. These parameters have been used successfully to estimate the target concentration in the clinic (25).

There is also ample evidence that interindividual differences in PD are sizeable and are associated with variables such as age, race and pathologies. Therefore in the future, effort should focus on resolving the population PD (e.g. with mixed effects approaches) hoping to eventually employ covariates for predicting individual PD characteristics. These studies must be designed in accordance with basic epidemiological principles, i.e. with populations where all the possible covariables (demographic or pathophysiological) can be completed for all subjects. The results from these PD studies together with their PK counterpart would constitute the basis for adequate use of concentrations and effects as therapeutic targets.

**II.5. Complex PK/PD situations and applicability of integrated models**

The concentration – effect relations can also be used to detect situations where there is an apparent lack of relation between concentrations and effect, when plasma concentrations can not be used directly as biomarkers. Such conditions exist, for example, when there is temporal disequilibrium between plasma and biophase, tolerance, presence of active metabolites, or enantiomers with distinct pharmacological activity, e.g. tramadol (26,27) and methadone (28). But even in these situations, the use of integrated PK/PD models could permit the prediction of the effect, reached after a certain drug dose or through a specific administration route, since the PD diagnoses the concentration necessary for a specific effect and the PK informs us of the dose corresponding to that concentration.

Since the kinetic and dynamic processes are intimately related to the temporal evolution of the pharmacological effect, combined PK/PD models have been developed in place of characterizing separately the concentration vs time and effect vs time relations.

Of all PK/PD models dealing with complicated conditions, the most developed and amply used is the effect compartment model. It resolves the possible disequilibrium between the central distribution compartment and the effect site via an empirical equilibration rate ke0 in a “link” model which allows the estimation of the concentrations at the effect site. In clinical practice, it is used for dose adjustment, particularly in anesthesia with propofol and phentanyll, since they are drugs with complex kinetics of multiple compartments posing the problem, discussed earlier, of choice of appropriate parameter for dose adjustment. For example, with propofol the effective concentration range for hypnosis is 2 - 3 mg/L. If, for
estimation of the therapeutic dose, we were to select the central V in a hypothetical patient of 70 kg weight with V = 10 L, the dose provided would be 30 mg, completely ineffective. If we were to select Vss for the same purpose (Vss = 466 L), the estimate would be 1428 mg, well above the therapeutic dose. The solution lies in performing the same task employing the peak-effect volume (Vpe = 20 L) estimated via the link model. The correct dose would then be 60 mg. This concept is actually programmed into the target controlled infusion pumps used in the operating room for anesthesia with propofol.

Another practical example of the importance of PK/PD integration is evident in a study where two oral formulations of ibuprofen were compared. The PK had been studied in healthy volunteers (typical study of bioequivalence) and the evolution of the effects (fever) in children with hyperthermia. At first, the kinetic study appeared to indicate that the two formulations were different in absorption as reflected in the time needed to assess Cmax (Tmax). Nevertheless, observation of the effect - time evolution in the children did not show any difference. Integrated PK/PD analysis of both populations jointly, with the use of NONMEM, allowed the determination of the causes of this discrepancy. Due to the nature of the indirect response mechanism, via which the fever process proceeds, the differences in the plasma concentration were not reflected in the observed therapeutic response. Eventually, the two formulations were bioequivalent (29).

The integration of PK and PD is key to understanding the use of TCI as an alternative to TDM. This latter approach, widely used at present, often fails precisely because it does not consider the pharmacological effect. The time seems ripe to start paying attention to the concentration effect relation and to think of strategies to individualize the dose with the help of the concentrations but without losing from sight the synthesis of the PK/PD concepts.

In conclusion, the appropriate combination of biomarker identification and selection, and bioanalytical methods for development and validation and the use of PK/PD models (including population approaches) for fitting data and predicting future clinical endpoints, can provide powerful insights and efficacious guidance for individual patients.

III. REFERENCES

Chapter 4. Pharmacogenetics and Pharmacogenomics

Kim Brøsen, M.D., Ph.D.

Institute of Public Health
Clinical Pharmacology
University of Southern Denmark
Winslowparken 19
DK-5000 Odense
DENMARK
Tel +45 65 50 37 51
Fax +45 65 91 60 89
Email kbrosen@health.sdu.dk
I. INTRODUCTORY REMARKS

It has been recognized for more than 50 years that genetic differences between people contribute to interindividual differences in the response to many commonly used drugs. Pharmacogenetics is the term used for more than 40 years to denote the science about how heritability affects the response to drugs. Pharmacogenomics is an apparently new science about how the systematic identification of all the human genes, their products, interindividual variation, intraindividual variation in expression and function over time may be used both to predict the right treatment in individual patients and to design new drugs. The term pharmacogenomics was coined in connection with the human genome project, but there is no internationally accepted consensus depicting any semantic differences between pharmacogenetics and pharmacogenomics, and in practice the two terms are used interchangeably. However it seems that most would use pharmacogenetics to depict the study of single genes and their effects on interindividual differences in drug metabolising enzymes, and pharmacogenomics to depict the study of not just single genes but the functions and interactions of all genes in the genome in the overall variability of drug response, whether this is caused by pharmacokinetics, pharmacodynamics or both.

The human genome is composed of 3.1 billion nucleotide bases, and the number of genes is about 26,000. Alternative splicing is relatively common and it may add to the complexity of the human proteome. The average gene consists of 3000 bases, but sizes vary greatly, with the largest known human gene being dystrophin at 2.4 million bases. The functions are unknown for over 50% of discovered genes. Another characteristic of the human genome is that by chance two unrelated subjects have approximately 99.9% DNA sequence homology. Still this leaves room for more than 1 million nucleotide bases being different between two randomly selected subjects. Very often the variability in DNA consists in only one nucleotide base change, and if this occurs in more than 100 subjects at a given position in the DNA, it is referred to as a single nucleotide polymorphism, abbreviated SNP and pronounced “snip”. SNPs make up about 90% of all human genetic variation and it occurs at every 100 to 300 bases. Two of every three SNPs involve the replacement of cytosine (C) with thymine (T), and SNPs can occur in both coding (gene) and noncoding regions of the genome. Many SNPs have no effect on cell, organ or whole body function, but some could predispose people to disease or influence their response to a drug.

In the classical pharmacogenetics a genetic polymorphism was a monogenic or Mendelian inherited pharmacological trait existing in at least two phenotypes (and presumably in at least two genotypes) the rarest of which exist in at least 1% of the population. The difference between the classical and the modern definition of a genetic polymorphism is among others, that a phenotype caused by homozygosity of an SNP with a frequency of 1% has a frequency in the population of one in 10,000.

Most of the classical pharmacogenetic entities have been discovered initially through an adverse drug event in a single or just a few subjects. This was followed by speculations into the putative pharmacological mechanism and subsequently into the identification of the target protein and eventually the target gene. By sequencing the target gene from subjects with different phenotypes functionally important SNPs could be discovered, and this would lead to a genotype test and eventually the application of genotyping in clinical practice. The modern pharmacogenomic approach is to reverse the sequence and originate from genotypes or haplotypes and subsequently to study if this in any way is related to differences in either efficacy or adverse drug events in individual subjects in regular clinical studies. To summarize, classical pharmacogenetics has searched the gene(s) for an abnormal drug response of proven clinical value whereas pharmacogenomics presently and in the future will search the bearing if any, on drug response of known genes, SNPs or haplotypes.

During the past 50 years pharmacogenetics has focussed on drug metabolizing enzymes, and there are a number of reasons why this is so. Firstly, drug metabolizing enzymes may vary more in function and expression than most other pharmacokinetic or pharmacodynamic targets, the reason being, that drug...
metabolising enzymes as a general rule are not per se related to health or survival. Secondly, provided that the drug is administered intravenously or absorbed completely following oral ingestion which is true for most but not all drugs) the area under the plasma concentration versus time curve (AUC) or the mean steady-state concentration of the drug is a very specific marker of drug elimination, ie. a low concentration means rapid elimination and a high plasma concentration means slow elimination. If in turn the elimination of drug depends on one enzyme, then the drug level is a very specific in vivo marker of the enzyme in question. Finally, the experimental setup, the drug assay technology and enzyme kinetics and pharmacokinetics applied in both in vitro and in vivo drug metabolism research is very much the same irrespective of what drug metabolizing enzyme or what drug is under study, and that creates a big flexibility in research.

While modern pharmacogenomics still is a science in the making, the pharmacogenetics of drug metabolising enzymes and particular the cytochrome P450 (CYP) enzymes has been in the focus for almost 30 years. This remainder of this chapter deals with in vivo methodologies and strategies in CYP pharmacogenetics.

II. CYP PHARMACOGENETICS

II.1. Cytochrome P450

Most drugs are lipophilic compounds that are mainly eliminated by oxidation catalyzed by the cytochrome P450 (CYP) enzyme system in the liver. The total number CYP genes in all species is 270, but the human CYP superfamily consists of 57 CYP genes and 33 pseudogenes organized in 18 families and 42 subfamilies. The CYP play a key role in the metabolism, and the endogenous substrates for CYP include fatty acids, eicosanoids, sterols and steroids, bile acids, vitamin D, retinoids and uroporphyrins. The CYP also represent the most important way for detoxification of many foreign chemical including drugs. The drug metabolizing CYP belong to families 1, 2 and 3.

II.2. CYP2D6

The CYP2D6 is the source of the sparteine/debrisoquine oxidation polymorphism, and in 7-9 % of Caucasian populations referred to as poor metabolizers the enzyme is not expressed due to mutations in the CYP2D6 gene on the long arm of chromosome 22. The frequency of poor metabolizers in Blacks and Orientals is only about 2-3%. CYP2D6 displays a marked allelic heterogeneity, and approximately 80 known variants, mainly in the form of single nucleotide polymorphisms in the gene, have been reported. However, the CYP2D6*3, *4 and *5 together predict about 90% of poor metabolizers. The CYP2D6 oxidizes approximately 60 drugs including all of the tricyclic antidepressants, some antipsychotics, selective serotonin reuptake inhibitors, antiarrhythmics, beta-adrenoceptor blockers and opiates.

Separation of individuals into the extensive and poor metabolizer phenotype can be done by either phenotyping or genotyping, and each approach has advantages and disadvantages. The classical model drugs for studying CYP2D6 are sparteine and debrisoquine, and they still are superior to all other probe drugs in this regard. The reason is that they both are almost exclusively oxidized by CYP2D6 and the amount of metabolite appearing in the urine reflects CYP2D6 very precisely, provided urine collection has been complete. The second important feature is that the fraction of the two drugs not metabolized is excreted unchanged in substantial amounts via the kidneys. Thus the metabolic ratios debrisoquine/4-hydroxydebrisoquine and sparteine/dehydrosparteines in 8-12 urinary samples following oral ingestion of 10 and 100 mg debrisoquine and sparteine, respectively, provide accurate and very specific measures of CYP2D6 irrespective of any urine loss. The distribution of both metabolic ratios is clearly bimodal, and extensive and poor metabolizers thus are clearly separated. Poor metabolizers have debrisoquine and sparteine metabolic ratios above 12.6 and 20 respectively (antimodes), and extensive metabolizers have
metabolic ratios below these values. However both sparteine and debrisoquine are obsolete drugs and hence no longer manufactured. Dextromethorphan is an alternative, and a dextromethorphan/dextrorphan ratio above 0.3 in most population studies has indicated the poor metabolizer phenotype. The problem with most substrates of CYP2D6 apart from sparteine and debrisoquine, in terms of serving as probe drugs, is that in the absence of CYP2D6 their preferred route of elimination is still oxidation via the same route to the same metabolite as that made by CYP2D6 only this being catalyzed by alternatively low affinity CYPs and at a much slower rate. The alternative CYPs also vary in activity, and although the parent compound/metabolite ratio in plasma or urine indeed reflects CYP2D6 very accurately, it does not separate extensive and poor metabolizers completely as do sparteine and debrisoquine metabolic ratios.

In Caucasians approximately 1% carries one or several extra copies of CYP2D6 and these individuals are always ultrarapid metabolizers. However only about 15% of phenotypically rapid metabolizers arbitrarily defined as subjects that have a metabolic ratio of sparteine below 0.15 have the duplicated allele. The molecular mechanism of ultrarapid metabolism in the remainder 85% is not known.

During the last 20 years the pharmaceutical industry has largely banned the development of CYP2D6 substrates because of the difficulties in handling a polymorphically metabolized drug. However there was a time where it was not common or possible to use in vitro methods (see chapter XX) to detect which CYPs catalyzed a particular drug including CYP2D6. There are three different in vivo methods that can be applied in order to find out if a drug is metabolized by CYP2D6. The phenotyped panel approach implies that the drug in question is given to 6-12 healthy extensive metabolizers and a similar number of poor metabolizers for either sparteine or debrisoquine. The drug under investigation is usually administered as a single oral dose, but sometimes it may be necessary to give it repeatedly in order to measure the steady-state concentration. CYP2D6 substrates are characterized by the fact that the AUC or Css is higher, usually 2-5 times higher in the poor metabolizers compared with the extensive metabolizers and that the total drug clearance is similarly lower in the former compared with the latter. Usually, but not always, the elimination half-life is 2-5 times longer in the poor compared with the extensive metabolizers. An alternative approach is to investigate a randomly selected group of typically 20-30 patients receiving the drug in question for treatment and to correlate their steady-state concentrations with the sparteine, debrisoquine or dextromethorphan metabolic ratios. For CYP2D6 the correlation is positive, i.e. the higher the steady-state concentration the higher the metabolic ratio. The third method is based on the use of a selective potent inhibitor and here quinidine has been the most commonly used. Six to twelve healthy extensive metabolizers take the drug either once or repeatedly and AUC or Css is determined before and during the concomitant intake of quinidine 100-200 mg/day. For a typical CYP2D6 substrate either of these pharmacokinetic parameters increases 2 to 5 times during quinidine. Quinidine is preferred over other inhibitors of CYP2D6 because it is the only one that selectively inhibits CYP2D6 and not other CYPs. The method can be refined by including a group of poor metabolizers as a negative control in whom the plasma concentration does not change during quinidine. Of the three methods, the quinidine inhibition method is the least commonly used because of the risk of proarhythmias caused by quinidine.

All three methods can be refined to look at individual pathways by detecting partial formation clearances of drugs in relation to the CYP2D6 phenotype. Research not only has identified most of the CYP2D6 substrates but it has also shown that CYP2D6 polymorphism displays marked dose dependent kinetics for most of its substrates and that it is the source of many important drug-drug interactions when two substrates that both are metabolized by CYP2D6 are co-administered.

II.3. CYP2C19

Approximately 20 years ago a genetic polymorphism in drug oxidation different from the sparteine/debrisoquine polymorphism was discovered through a bimodal distribution of the aromatic 4-hydroxylation of the (S)-enantiomer of the antiepileptic drug mephenytoin. In Caucasians 2-3% are poor
metabolizers. The S-mephenytoin oxidation polymorphism displays marked interethnic variability as 15-20 % of Orientals are poor metabolizers. About 10 years ago it was reported that the CYP2C19 is the source of the S-mephenytoin oxidation polymorphism. CYP2C19 also displays a considerable allelic heterogeneity, and so far 9 different single nucleotide polymorphisms have been reported in the poor metabolizers. However, the CYP2C19*2 and the CYP2C19*3 mutations reflecting G→A SNPs in exon 5 and 4 respectively still account for approximately 90 % of the poor metabolizers. The CYP2C19 is an important mediator of the biotransformation of tertiary amine tricyclic antidepressants (N-demethylation of amitriptyline, clomipramine, imipramine and trimipramine), all of the proton pump inhibitors, the bioactivation of proguanil, moclobemide and several other drugs.

Mephenytoin is the classical model drug used for phenotyping, but the drug is no longer in clinical use and hence it is no longer manufactured. However it is still available for pharmacogenetic research. Following the ingestion of a single oral dose of mephenytoin, urine should be collected for up to 8 or 12 hours. Urine is analysed for (R)- and (S)-mephenytoin, and the ratio between the chromatographic peak areas, the S/R ratio, is determined. The stereoselective metabolism of mephenytoin is abolished in the poor metabolizers and hence the S/R ratio is close to unity. In extensive metabolizers, (S)-mephenytoin is rapidly hydroxylated and (R)-mephenytoin is slowly demethylated. Thus the S/R ratio is less than one in extensive metabolizers and usually is less than 0.1. An acid labile metabolite is formed in extensive metabolizers and this is gradually converted to (S)-mephenytoin even if the urine is kept at -20°C. Thus the longer the sample is kept before the assay the higher the S/R ratio. If the urine sample is treated with acid then the S/R ratio becomes much higher in extensive metabolizers but it does not change in the poor metabolizers. In all samples where the initial S/R ratio is determined to be above 0.5 it is advisable to add acid and repeat the assay. In the poor metabolizers the S/R ratio does not change whereas in the extensive metabolizers it increases by a factor 10 or more.

There is certainly a need for a better model drug than mephenytoin, and omeprazole appears to be a candidate. Following a single oral dose of 20 mg of omeprazole a blood sample is drawn after 3 hours. The omeprazole/5-hydroxyomeprazole ratio in plasma has been reported to be above 7 in the poor metabolizers and below 5 in the extensive metabolizers.

The in vivo methods for determining if a drug is metabolized are not different from what has been described above for CYP2D6 except that the inhibitor method is not used. Drugs such as fluvoxamine and moclobemide are potent inhibitors of CYP2C19 but they are not selective for this CYP. Thus an increase in plasma concentration of a drug during concomitant intake of either of these drugs does not prove that the drug in question is metabolized by CYP2C19.

II.4. CYP2C9

CYP2C9 is a major enzyme catalyzing the biotransformation of warfarin, phenytoin, fluvastatin, several NSAIDs, tolbutamide and other oral antidiabetics. The CYP2C9 also is the source of a genetic polymorphism but contrary to the CYP2D6 and CYP2C19 polymorphism this was not discovered through a bimodal distribution of a metabolic ratio for one of the drugs metabolized by the enzyme. Rather it was discovered by sequencing of the gene and detection of several SNPs.

CYP2C9*1 is the wild-type allele, and besides there are two important single nucleotide polymorphisms the CYP2C9*2 associated with a functionally important Arg144Cys substitution and the CYP2C9*3 associated with another important Ile359Leu substitution. In Caucasians the frequencies of the 6 different genotypes are 65-70 %, 15-20 % and 8-10 % for CYP2C9*1/*1, CYP2C9*1/*2 and CYP2C9*1/*3. The poor metabolizer genotypes CYP2C9*2/*2, CYP2C9*3/*3 and CYP2C9*2/*3 each occur in about 1-2%.
Tolbutamide has been proposed to be a candidate for a model drug to probe for CYP2C9, but it is not possible to separate the 6 common genotypes completely. In one clinical study by Scordo et al. (1) the average maintenance dose of warfarin [mg/week (±s.d.)] to achieve the desired International Normalized Ratio (INR) (a measure of anticoagulant effect) was: CYP2C9*1/*1: 39 (±15), CYP2C9*1/*2: 28 (±11), CYP2C9*1/*3: 21 (±7), CYP2C9*2/*2: 21 (±4), CYP2C9*2/*3: 18 (±9) and CYP2C9*3/*3: 9 (±4). Although this study differs in that it does not measure the pharmacokinetics but rather the dose required in different genotypes to achieve a desired pharmacodynamic end point, it is still a paradigm for studying the clinical relevance of pharmacogenetics in drug metabolism. It clearly shows that the CYP2C9 oxidation polymorphism is an important determinant of the warfarin dose, but it also shows that within each geno-(pheno-)type there is a considerable variability caused by other genes, the constitution of the patient and the environment. CYP2C9 genotyping before treatment with warfarin probably has limited, if any, value in practice.

II.5. Pheno- or genotyping for drug metabolizing enzymes

Pheno- or genotyping as an aid for choosing the optimal dose from the start in theory should be performed if the drug is exclusively metabolized by a single CYP, if the drug has a low therapeutic index and if clinical dose titration is not feasible. For CYP2D6 the possible candidates include tricyclic antidepressants, some antipsychotics, and some antiarrhythmics and for CYP2C9 the possible candidates could be warfarin and phenytoin. However pheno- or genotyping for CYP enzymes has never achieved widespread use in clinical practice because, as explained above, the response is not determined alone by a single enzyme or gene.

The advantage of phenotyping compared with genotyping is, that phenotyping gives an up-to-the-minute account of the CYP in question integrating the result of the genetic, constitutional and environmental influences. The disadvantage is that it involves the ingestion of a model drug which is often obsolete as a therapeutic agent and hence no longer manufactured and also that it is necessary to collect urine or to draw a blood sample. Genotyping separates patients into categories, it is not influenced by environmental or constitutional factors and it only has to be performed once in the life time of the patient.

III. FUTURE DIRECTIONS

On the basis of the last 25 years of intensive studying the pharmacogenomics/-genetics of CYP enzymes a number of general statements regarding the role of genetic factors for variation in drug response can probably be formulated. First until proven otherwise the response to any drug is always determined to some extent by genetic factors. However drug response is never determined by a single gene or by a group of genes alone. It is rather determined by several interacting genes and with important influences from the environment and from the constitution of the patient.

Now that the human genome is known in practically all detail there is widespread optimism, that in the near future, it will be possible to tailor the treatment to the individual patient on the basis of the patient’s genotype. This author does not entirely share this optimism. Genotyping as an aid to select the right dose of the right drug in the individual patient would be of theoretical use if the response is mainly determined by a single gene or a limited group of genes, and if all of the environmental and constitutional influences have a more limited influence, and besides are known in detail and can be measured in the individual patient. It is anticipated that less than 10 % of drugs in 10 years from now will be prescribed following a pharmacogenetic test. However by extrapolation from the 25 or more years of experience in the CYP field studying pharmacogenetics/-genomics will lead to new important insights and discoveries that will ultimately lead to the development of new and better drugs and to the rational use of drugs that are already on the market. According to this author’s view it is here that the true importance and benefits of pharmacogenetics lies.
IV. REFERENCES


This chapter was written on the basis of numerous studies that have not been cited in the text. As a textbook that gives a comprehensive overview and is reasonably well updated the following can be recommended:


There is a website with a constant update and comprehensive overview:

Chapter 5. Paediatric Drug Research

Imti Choonara, M.D.

Professor in Child Health
Academic Division of Child Health (University of Nottingham)
Derbyshire Children’s Hospital
Derby DE22 3DT
UK
Tel 01332 724693
Fax 01332 724697
Email imti.choonara@nottingham.ac.uk
I. INTRODUCTORY REMARKS

Studies in Europe (1) and Australia (2) have shown that a significant number of children, both in hospital and the community, receive unlicensed and off-label medicines. Unlicensed medicines are where the medicine has been modified from that specified in its product license (3). This may involve crushing a tablet in order to make it into a suspension so that a young child can take the medicine orally. Off-label medicines are those that are used in a different manner than recommended in the product license. This may involve using the medicine in a different age group, for a different indication, at a different dose or by a different route to that recommended (3).

II. PAEDIATRIC DRUG RESEARCH

II.1. Why?

The licensing process was introduced in response to drug toxicity that affected children and adults. In particular, the Medicines Act (UK 1968) and the Kefauver-Harris amendment (USA 1962) were passed following 1) drug toxicity in the newborn infant – e.g. the grey baby syndrome due to chloramphenicol and 2) drug toxicity in the developing fetus – e.g. phocomelia due to thalidomide. It is ironic that legislation introduced following drug toxicity in the newborn infant and the developing fetus has failed to ensure that medicines used in paediatric patients are fully tested in relation to efficacy and toxicity. The use of unlicensed and off label medicines is thought to be associated with a greater risk of drug toxicity (4).

An additional problem associated with the widespread use of off label medicines is the lack of appropriate formulations for young children. Young children cannot swallow tablets and they therefore require liquid formulations. Recent studies have shown that many young children are prescribed tablets or capsules even though they are too young to swallow them (5).

Drug toxicity in children is different to that in adults (6). One therefore needs to study medicines in paediatric patients in order to prevent future cases of drug toxicity. Different mechanisms of drug toxicity in children are illustrated in Table 1.

II.2. Who?

Paediatric drug research needs to involve the pharmaceutical industry working in partnership with paediatric health professionals. The latter group consists of doctors, pharmacists and nurses with paediatric expertise. Ideally paediatric clinical pharmacologists, who have both the expertise of other paediatric health professionals and an understanding of clinical pharmacology should be involved, especially in relation to the design of the clinical trials.

The pharmaceutical industry has previously been reluctant to become involved in drug research in children. The legislative changes that have been introduced in the USA (the FDAMA and the Pediatric Rule) have provided a significant financial incentive to study medicines (7). Discussions are currently taking place in Europe with regards to introducing some financial incentive for the European pharmaceutical industry.

<table>
<thead>
<tr>
<th>Year</th>
<th>Drug / Compound</th>
<th>Age group</th>
<th>ADR</th>
<th>Mechanism</th>
</tr>
</thead>
<tbody>
<tr>
<td>1886</td>
<td>Aniline dye</td>
<td>Neonates</td>
<td>Methaemoglobinemia</td>
<td>Percutaneous absorption</td>
</tr>
<tr>
<td>1956</td>
<td>Sulphisoxazole</td>
<td>Neonates</td>
<td>Kernicterus</td>
<td>Protein displacing effect on bilirubin</td>
</tr>
<tr>
<td>Year</td>
<td>Medication</td>
<td>Age Category</td>
<td>Adverse Effect</td>
<td>Reason</td>
</tr>
<tr>
<td>------</td>
<td>------------------</td>
<td>-----------------------</td>
<td>-------------------------</td>
<td>-------------------------------</td>
</tr>
<tr>
<td>1959</td>
<td>Chloramphenicol</td>
<td>Neonates</td>
<td>Grey baby syndrome</td>
<td>Impaired metabolism</td>
</tr>
<tr>
<td>1979</td>
<td>Sodium valproate</td>
<td>Young children (&lt; 3 years)</td>
<td>Hepatic failure</td>
<td>Abnormal metabolism?</td>
</tr>
<tr>
<td>1980</td>
<td>Salicylate</td>
<td>Children</td>
<td>Reye's syndrome</td>
<td>Unknown</td>
</tr>
<tr>
<td>1990</td>
<td>Propofol</td>
<td>Children</td>
<td>Metabolic acidosis</td>
<td>Unknown</td>
</tr>
<tr>
<td>1996</td>
<td>Lamotrigine</td>
<td>Children</td>
<td>Skin reactions</td>
<td>Unknown</td>
</tr>
</tbody>
</table>

(Reproduced with permission from Paed Perinatal Drug Ther (6))

II.3. Where?

One cannot perform a paediatric clinical trial in an adult clinical trials unit. It is accepted that sick children need to be treated by paediatric health professionals. Similarly for clinical trials involving children, paediatric health care professionals need to be involved, ideally within a paediatric unit. It is important to recognise that clinical trials and other aspects of paediatric drug research can be performed in district general hospitals (8). For general paediatric conditions, these units are probably preferable to tertiary centres where one is more likely to see a highly selective patient group that is not representative of children throughout the community.

For those clinical trials that involve children as outpatients, it is important that the children are assessed in a child friendly location, i.e. safe with toys and play therapists available. It may also be appropriate to assess the child at home.

II.4. Which Medicines?

The success of any clinical trial is related to the clinical need for the medicine. Investigators are more likely to participate in a study of a medicine which is likely to result in significant clinical benefit to children than one where there is already satisfactory treatment. The clinical need of the medicine will also be taken into account by the ethics committee. Ethics committees are more likely to recognise that a clinical trial is appropriate in children if there is no current treatment available. This does not mean, however, that the clinical trial will automatically be approved as the design of the study may be inappropriate.

The International Conference on Harmonisation, which includes representatives from the European Medicines Evaluation Agency (EMEA), the Food and Drug Administration (FDA) and Japan have issued ICH E11, Clinical Investigation of Medicinal Products in the Paediatric Population (7). This guidance categorises medicinal products and their value in children into three categories. These are shown in Table 2. The aim is to encourage the study of medicines in the first two groups where there is the greatest potential clinical benefit.

<table>
<thead>
<tr>
<th>Table 2</th>
<th>ICH Classification of medicinal products for children</th>
</tr>
</thead>
<tbody>
<tr>
<td>Medicinal products for diseases predominantly or exclusively affecting paediatric patients</td>
<td></td>
</tr>
<tr>
<td>Medicinal products intended to treat serious or life threatening diseases occurring in both adults and paediatric patients for which there are currently no or limited therapeutic options</td>
<td></td>
</tr>
<tr>
<td>Medicinal products intended to treat other diseases and conditions</td>
<td></td>
</tr>
</tbody>
</table>
II.5. When?

The timing of studies in children is clearly dependent upon several factors. These include whether one is dealing with a serious or life-threatening disease for which there is currently no or limited treatment available. In this situation early paediatric studies are essential. However, where existing treatment is available then clinical trials in children should only be conducted after initial safety data has been established in adults.

II.6. Which paediatric patients?

The clinical nature of the medicine will determine whether it needs to be studied in neonates, infants or children. It is important that investigators are aware of the ICH Guidance in relation to the classification of the five different age groups in relation to paediatric patients. These are shown in Table 3.

<table>
<thead>
<tr>
<th>Category</th>
<th>Age</th>
</tr>
</thead>
<tbody>
<tr>
<td>Preterm neonates</td>
<td>&lt;36 weeks gestation, 0 – 27 days</td>
</tr>
<tr>
<td>Full-term neonates</td>
<td>0 – 27 days</td>
</tr>
<tr>
<td>Infants and toddlers</td>
<td>28 days – 23 months</td>
</tr>
<tr>
<td>Children</td>
<td>2 – 11 years</td>
</tr>
<tr>
<td>Adolescents</td>
<td>12 – 17 years</td>
</tr>
</tbody>
</table>

II.7. How?

II.7.A. Study design

The study design is crucial in that a poorly designed study will fail to attract investigators, obtain ethical approval and recruit children. Investigators need to ask the following questions:

a. Which paediatric age group is most likely to benefit from the medicine?

b. Should a placebo be included in the trial design? (Placebo is appropriate if there is no existing treatment for the condition. If however, effective therapy is available, then the use of a placebo is neither appropriate nor ethical.)

c. How will the pharmacodynamic effect be studied?

d. Which pharmacokinetic parameters, if any, need to be determined?

e. What is the likelihood of significant drug toxicity?

Regulatory authorities are more likely to raise questions about clinical trials in children than in adults (9). The duration of clinical trials in adult and paediatric patients is similar (in Finland two thirds completed within 12 months) (9). Clinical trials in healthy adult volunteers, however, are significantly shorter (over two thirds completed within 6 months) (9).

II.7.B. Pharmacokinetics

It is important to ensure that the minimum number of samples, involving the smallest amount of blood possible is collected from each patient. Microassays may need to be developed to measure drug concentrations in small volumes of blood. Information regarding the metabolic pathway and pharmacokinetic parameters in adults is essential before commencing pharmacokinetic studies in paediatric patients.

The use of population pharmacokinetics whereby a larger number of children are involved but fewer samples are collected from each patient should be considered (10). It is usually appropriate to carry out pharmacokinetic studies in a subgroup of the children recruited for the clinical trial. It should not be made
a precondition for entry into the clinical trial as this may result in the loss of a significant number of
cchildren from the study.

II.7.C. Non-invasive methods
Consideration needs to be given to the use of non-invasive techniques such as the caffeine breath test. The
caffeine breath test has been used as a probe for CYP1A2 enzyme activity (11). It involves the use of a
stable isotope of caffeine and the collection of breath samples for two hours after administration of the
caffeine. It has been used to study drug interactions (induction and inhibition) and also the effect of
disease on drug metabolism11.

II.7.D. Pharmacodynamics
It is often difficult to study pharmacodynamic effect in younger patients. For certain conditions,
measuring the pharmacodynamic effect is not difficult, e.g. seizures in patients with epilepsy (12),
mortality in children with leukaemia. For other conditions, however, it is more difficult to assess
pharmacodynamic effect, e.g. bronchodilators in infants under the age of 18 months, assessing pain relief
in pre-verbal children and neonates. There are validated pain scales appropriate for use in paediatric
patients of different ages (13). It is, therefore, essential that one uses a validated pain scale if one is
studying an analgesic drug.

II.7.E. Pharmacovigilance
Drug toxicity in children is different to that in adults (6). This may be due to impaired drug metabolism or
altered protein binding, but may also be idiosyncratic. As the child is developing they may be prone to
different toxicities to adults. The principles of pharmacovigilance in children are similar to that in adults.
Consideration should be given to setting up an Independent Safety Monitoring Board if there is the
potential for significant toxicity.

III. CONCLUSIONS
Paediatric drug research is more difficult than similar studies in adults. Paediatric drug research involves
patients whereas many adult studies involve volunteers. It is up to paediatric health professionals and the
pharmaceutical industry to work together to ensure that we study the right medicines with an appropriate
design to ensure that children receive medicines that are fully evaluated scientifically. Such an approach
will increase efficacy and hopefully reduce toxicity.

IV. REFERENCES
Ther 1997;1:52-55.
4. Turner S, Nunn AJ, Fielding K, Choonara I. Adverse drug reactions to unlicensed and off-label
5. Schirm E, Tobi H, de Vries TW, Choonara I, De Jong-van den Berg LTW. Lack of appropriate
6. Choonara I, Rieder MJ. Drug toxicity and adverse drug reactions in children – a brief historical
7. Spielberg SP. Paediatric therapeutics in the USA and internationally: an unparalleled opportunity.
Chapter 6. Phase I Studies (Human Pharmacology)

Edward M. Sellers, M.D., Ph.D. ¹

Patrick du Souich, M.D., Ph.D. ²

¹ Ventana Clinical Research Corporation
and
Departments of Pharmacology, Medicine and Psychiatry
University of Toronto
Toronto, Ontario
CANADA

² Département de pharmacologie
Faculté de médecine
Université de Montréal
Montréal, Québec H3C 3J7
CANADA
I. INTRODUCTORY REMARKS

I.1 Objective and overview of Phase I studies

Phase I studies are designed to determine the metabolic and pharmacological actions of the drug in humans, the side effects associated with increasing doses, and if possible, to gain early evidence on effectiveness.

Phase I studies are usually conducted in healthy volunteer subjects, but may be conducted in patients when information on effectiveness is not possible in healthy volunteers or when adverse drug effects are potentially important. The total number of subjects included varies with the drug, but is generally in the range of 20 to 80. Being the first exposure of an investigational new drug to humans, the subjects are closely monitored.

Phase I studies can also evaluate drug kinetics and metabolism, and the mechanism of action in humans. The Phase I study must provide sufficient information about the drug's pharmacokinetics and pharmacological effects so to allow the design of Phase II studies.

II. REQUIRED PRE-CLINICAL SAFETY EVALUATION

Pre-clinical safety evaluation information is required to estimate an initial safe starting dose for the human trials and the identification of parameters for clinical monitoring for potential adverse effects. The pre-clinical safety studies should be sufficient to characterize potential toxic effects under the conditions of the proposed clinical trial. The following aspects of the drug should be evaluated prior to human exposure (1):

II.1. Safety pharmacology
Safety pharmacology aims to assess the effects of the drug on vital functions, such as cardiovascular, central nervous and respiratory systems, of the animal.

II.2. Toxicokinetic and pharmacokinetic studies
Information on absorption, distribution, metabolism and excretion (ADME) in animals should be made available to compare human and animal metabolic pathways.

II.3. Single dose toxicity studies
The single dose (acute) toxicity of a drug has to be evaluated in two mammalian species prior to the first human exposure. A dose escalation study is considered an acceptable alternative to the single dose design.

II.4. Repeated dose toxicity studies
A repeated dose toxicity study has to be determined in two species (one non-rodent). The duration of repeated dose toxicity studies is related to the duration, therapeutic indication and scale of the proposed clinical trial. In principle, the duration of the animal toxicity studies should be equal to or exceed the duration of the human clinical trials. A repeated dose toxicity study for a minimum duration of 2-4 weeks would support Phase I (Human Pharmacology) studies up to 2 weeks in duration. Beyond this, 1-, 3-, or 6-month toxicity studies would support Phase I human clinical trials for up to 1, 3, or 6 months, respectively (1).

II.5. Other studies required prior to human exposure
Local tolerance should be studied in animals using routes relevant to the proposed clinical administration. Such studies are particularly important for parenteral products.
Prior to first human exposure, in vitro genotoxicity studies for the evaluation of mutations and chromosomal damage are generally needed. However, completed carcinogenicity studies are not usually needed in advance of the conduct of clinical trials unless there is cause for concern (1).

Reproduction toxicity studies should be conducted in accordance to the population that is to be exposed to the drug. There are regional differences in the timing of reproduction toxicity studies to support the inclusion of women of childbearing potential in clinical trials. In Japan, assessment of female fertility and embryo-foetal development should be completed prior to the inclusion of women of childbearing potential using birth control in any type of clinical trial. In the European Union, assessment of embryo-foetal development should be completed prior to Phase I trials in women of childbearing potential and female fertility studies prior to Phase III trials. In the United States, women of childbearing potential may be included in early, carefully monitored studies without reproduction toxicity studies provided appropriate contraception and precautions are taken to minimize risk (1).

III. EXPERIMENTAL DESIGN

Phase I studies usually have non-therapeutic objectives and as such are conducted in healthy volunteer subjects. The studies may be open, baseline controlled or may use randomisation and blinding. The design depends upon the aim of the study. When the objective is to determine the tolerability of the dose range expected to be used in later clinical studies and to assess the nature of adverse reactions, the study will typically include both single and multiple dose administration and a placebo arm. The protocol should specify the toxicity-based discontinuation of the study, as well as the dose adjustment rules in presence of toxicity (2).

In Phase I studies, pharmacokinetic studies aim to assess the drug’s half-life and clearance to anticipate accumulation of parent drug or metabolites. Pharmacokinetics may be determined as a part of efficacy, safety and tolerance studies or via separate studies. Depending upon the drug, pharmacodynamic studies may be conducted in healthy volunteer subjects or in patients with the target disease. Drug blood concentrations can be related to response (PK/PD studies). Preliminary studies of activity or potential therapeutic benefit may be conducted in Phase I as a secondary objective when the study is conducted in patients (2).

IV. DETERMINATION OF THE MAXIMUM RECOMMENDED STARTING DOSE (MRSD)

The MRSD for the first human clinical trial is derived from the no-observed adverse effect levels (NOAELs) in the most appropriate species, conversion of NOAELs to human equivalent doses (HED), and application of a safety factor. Pre-clinical toxicology studies will generate basically three types of findings that can be used to determine the NOAEL: a) overt toxicity, e.g., clinical signs, macro- and microscopic lesions, b) surrogate markers of toxicity, e.g., serum liver enzyme levels, and c) exaggerated pharmacodynamic effects (3).

The HED shall be estimated with the NOAEL estimated in the most relevant species. Alternatively, in the absence of data on species relevance, the most appropriate species for deriving the MRSD is the most sensitive species, e.g., the species with the lowest HED. Several factors could influence the choice of the most appropriate species including (3):

a) differences in ADME of the drug between the species;
b) evidence indicating that a particular model is predictive of human toxicity; and
c) limited biological cross-species pharmacologic reactivity of the therapeutic.
The NOAEL for medicaments administered systemically to animals is accurately extrapolated to other species and human when doses are normalized to body surface area (mg/m²). The easiest approach is using the following equation (3):

\[
HED = \text{animal NOAEL} \times \left(\frac{W_{\text{animal}}}{W_{\text{human}}}\right)^{1-b}
\]

Where \(W\) is the weight in kg and \(b\) (equal to 0.67) is a correction factor used to convert mg/kg to mg/m².

For instance, if the NOAEL of an investigational drug estimated in the relevant species (monkey of 3 kg) is 50 mg/kg, the HED will be:

\[
HED = 50 \text{ mg/kg} \times (3 \text{ kg}/65 \text{ kg})^{0.33} = 50 \text{ mg/kg} \times 0.3624 = 18 \text{ mg/kg}
\]

Once the HED has been determined, a safety factor is applied to provide a margin of safety. This safety factor allows for variability in extrapolating from animal toxicity studies to studies in humans resulting from (3):

a) uncertainties due to enhanced sensitivity to therapeutic activity in humans versus animals,
b) difficulties in detecting certain toxicities in animals,
c) differences in receptor densities or affinities,
d) unexpected toxicities, and
e) interspecies differences in ADME.

In practice, the MRSD for the clinical trial is determined by dividing the HED by a default safety factor of 10.

Under selected circumstances, a safety factor greater than 10 should be used (3):

a) presence of steep dose-response curve for important toxicities in the most relevant species or in multiple species,
b) severe toxicity or damage to an organ system,
c) irreversible toxicity in animals,
d) non-monitorable toxicity, such as histopathologic changes in animals that are not readily monitored by clinical pathology markers,
e) presence of significant toxicities without prodromal indicators,
f) non-predictable and unexplained mortality,
g) variable bioavailability between species, with poor bioavailability in the test species used to derive the HED,
h) large variability in doses or AUC levels eliciting a toxic effect,
i) questionable study design or conduct, such as few dose levels, wide dosing intervals, or large differences in responses between animals within dosing groups,
j) novel therapeutic targets, and
k) use of animal models with limited utility to evaluate a medicine because of very limited interspecies cross-reactivity or pronounced immunogenicity, or because its effect is elicited by mechanisms that are not known to be conserved between animals and humans.

For selected medicines where an exhaustive and rigorous toxicologic testing has been made, several circumstances allow the use of safety factors of less than 10 (3):
a) the medicament is a member of a well-characterized class, has a similar metabolic profile and bioavailability, presents similar toxicity across all the species tested including humans, and the drug is administered by the same route, schedule, and duration of administration,

b) the toxicity elicited by the drug is easily monitored, is reversible and predictable, and exhibits a moderate to shallow dose-response relationship with toxicities that are consistent across the tested species, and

c) the NOAEL is estimated based on toxicity studies of longer duration than required for the proposed clinical schedule in healthy volunteers.

V. STUDY POPULATION AND SAMPLE SIZE

Phase I studies are usually conducted in healthy volunteer, male and females older than 18 years. Patients may be included when information on effectiveness is desired and it is impossible to assess in healthy volunteers or when the information obtained is difficult to extrapolate to a patient population, e.g. antibacterials, drugs used in affective disorders, antipsychotic drugs, analgesic drugs, antihypertensives, or when drug adverse effects are potentially important, such as in cancer chemotherapy, anti-HIV drugs, etc (2).

The description of the study population should identify the characteristics that are important to understand to interpret and apply the study results. The description of the study population should identify important inclusion and exclusion criteria, demographic characteristics, baseline values of any clinically relevant variables that would be important to understand the treatment effect, and other characteristics of the population that have implications for the extent to which results can be generalised. The selection of the population, e.g. the inclusion and exclusion criteria, is defined according to the population to be included, e.g. healthy volunteer or patient. The total number of subjects included in Phase I studies varies with the drug, but is generally in the range of 20 to 80 (2).

VI. ROUTE OF ADMINISTRATION AND DURATION OF THE STUDIES

The route of administration used of the study medicine during Phase I studies is usually the intended route for the commercialised drug. Ideally, the duration of drug administration should be equal to the intended duration of drug treatment. For drugs intended for chronic use, this is not feasible and the duration of administration shall be determined based upon the kinetics or the dynamics of the drug, e.g. until steady state is reached or until the maximal effect is achieved.

VII. MONITORING

To assess safety, monitoring of the subjects enrolled in the Phase I study is of utmost importance, and its extent will depend upon the aim of the study. Vital signs, blood chemistry and ECG should be monitored. In addition, depending upon the characteristics of the toxicity of the drug, other parameters of organ function shall be monitored. Whenever the Phase I study aims also to assess effectiveness, drug response will be monitored.

Monitoring is critical to decide when dose adjustment should be done or when the study should be discontinued due to unacceptable toxicity (2).

VIII. EXAMPLE OF A PHASE I EXPERIMENTAL PROTOCOL

The investigational drug XXX is a product that in clinical practice is intended to be administered for four consecutive days. Two Phase I studies were designed, the first one, where the volunteers received a single escalating dose and the second, included repeated dosages for four days.
VIII.1. Study Objective(s)

VIII.1.A. Single escalating dose study
Primary: To assess the safety of single doses of XXX.
Secondary: To gather preliminary information about the subjective, performance-altering, and effects of single doses of XXX.

VIII.1.B. Repeated i.m. doses study
Primary: To determine the maximum tolerated sub-chronic (4 days) daily dose of XXX.
Secondary: To determine the minimum number of individual doses into which that maximum tolerated sub-chronic daily dose of XXX must be divided, and to test whether the dose and schedule so determined can also be tolerated during a treatment period of seven days.

VIII.2. Study Design

VIII.2.A. Single escalating i.m. doses of XXX
There will be a total of up to 72 participants divided into 9 groups, with 8 subjects in each group, 6 of who will receive active drug and 2 of whom will receive placebo (the vehicle in which XXX is normally diluted). The active drug assignment within groups will be randomised and double-blind. Groups receiving higher doses will do so only after the previous group has received the next lowest dose (i.e. in an ascending manner). Each subject will receive only one injection. The study will be terminated at any dose which produces clinically significant adverse effects. The study will last 21 days.

VIII.2.B. Repeated i.m. doses of XXX
This is a randomised, double-blinded, placebo-controlled, single-site tolerance study of XXX. Up to 56 normal, healthy males or females, aged 18 to 45 years will be included in the study. Subjects will be divided into seven groups of eight. The duration of treatment is four inpatient days for the first six groups and seven inpatient days for the last group. After each injection of XXX or placebo is administered, the drug’s objective and subjective effects are monitored.

VIII.3. Study Population
Healthy males or females, ages 18 to 50.

VIII.4. Inclusion Criteria
A subject will be eligible only if all of the following criteria apply:

a. Males or females between the ages of 18 and 50.
b. No clinically important abnormal physical findings at the screening examination.
c. Normal or clinically acceptable ECG.
d. Normal blood pressure (systolic: 90-140 mmHg; diastolic: 50-90 mmHg) and heart rate (40-100 bpm).
e. Body Mass Index of 19.0-29.0 (kg/m²).
f. Ability to communicate well with the investigator and to comply with the requirements of the entire study.
g. Willingness to give written informed consent (prior to any study-related procedures being performed) and to be able to adhere to the study restrictions and examination schedule.
VIII.5. Exclusion Criteria

A subject will not be eligible if any of the following criteria apply:

a. Administration of any investigational drug in the period 0 to 45 days before entry to the study.
b. Use of any prescription medication during the period 0 to 30 days or over-the-counter medication during the 0 to 5 days before entry to the study.
c. Donation or loss of greater than 400 ml of blood in the period 0 to 12 weeks before entry to the study.
d. Serious adverse reaction or hypersensitivity to any drug.
e. Inability to communicate or co-operate with the investigator because of a language problem, poor mental development or impaired cerebral function.
f. History of drug dependence (except tobacco) or psychiatric illness within the past 2 years.
g. Consumption of alcohol within 24 hours prior to dose administration.
h. Females who are lactating or at risk of pregnancy (i.e. sexually active with fertile males and not using an adequate form of birth control).
i. Females with a positive serum pregnancy test at screening or positive urine pregnancy test on admission to study site.
j. Presence of pain incurred by unknown causes.
k. History of asthma or other respiratory disease.
l. History of neurologic or neuromuscular disease.
m. History of hypotension or cardiovascular disease.
n. History of bladder or urethral disease.
o. Positive urine drug screen for drugs with a high potential for abuse and low persistence in the urine.
p. Inability to refrain from smoking during study days.
q. Any other condition which, in the opinion of the investigators, is likely to interfere with the successful collection of the measures required for the study.

VIII.6. Other Study Eligibility Criteria Considerations

All detailed information regarding warnings, precautions, contraindications, adverse events, and other significant data pertaining to the study drug being used in this study will be taken into account, whether or not explicitly mentioned in this protocol. Such documents may include, but are not limited to, the Investigator's brochure or equivalent document provided by the Sponsor.

VIII.7. Study Drugs And Dosages

VIII.1. Single escalating i.m. doses of XXX

The initial starting dose was chosen as 1/20 of the chronic non-toxic dose in Rhesus monkeys (NOAEL = 1 µg/kg). Subjects will be admitted to the overnight facilities of YYY CRC the night prior to the study. On the day of the study, drug or placebo will be injected at 8 AM. Safety, physiologic, subjective, performance and effect measures will be collected as outlined for the next 24 hours. At 8 AM the day following the study day, a medical assessment, laboratory tests and ECG will be conducted prior to subject discharge. Subjects will be reassessed 48 hours after discharge for any late onset adverse events. Figure 1 illustrates the experimental design.

VIII.2. Repeated i.m. doses of XXX

The duration of treatment is four inpatient days for the first six groups and seven inpatient days for the last group. The total daily dose of XXX will either be increased or decreased and the number of doses into
which that total daily dose is divided will either remain the same or be diminished (resulting in larger individual doses), both based on tolerance to previously administered dose schedules. The first group will receive the maximal single dose used (48 µg/day) of XXX i.m divided into four equal portions administered at 4-hour intervals, for up to 4 days. The experimental design is illustrated in figure 2.

VIII.8. MISCELLANEOUS

The protocol contains information about aspects such as overdose with XXX and toxicity management, concurrent medications and non-drug therapies, XXX management, measurements and evaluations, premature discontinuation, data analysis methods, adverse effects and study administration.

IX. REFERENCES

Figure I. Nine groups of 8 healthy volunteers (6 receiving XXX and 2 placebo) not tolerated at 45 ng/dose.
Figure 2. Randomized, double-blind, placebo-controlled study of multiple-dose tolerance of i.m. XXX in healthy volunteers.
Chapter 7. Follow-Up of Drugs After Market Entry

Jacques LeLorier, M.D., Ph.D., FRCPC

Centre de recherche - Hôtel-Dieu
Centre hospitalier de l'Université de Montréal
Pavillon Masson
3850 St-Urbain
Montréal, Québec H2W 1T8
CANADA
I. INTRODUCTORY REMARKS

There is a general agreement with the fact that our operational knowledge on drugs at the time they enter the market is grossly inadequate. It is also generally accepted that, in most cases, delaying entry into the market by requesting additional animal or clinical studies would not answer the remaining questions and would only delay the patients’ access to useful and sometimes life-saving drugs (1). The solution is therefore to continue studying drugs in a formal way for an indeterminate period of time following their entry into the market (2). Those who consider indeterminate too long a period of time might wish to recall the cisapride experience.

Although few active participants and observers of the medication scene would disagree with the above statements, there is considerable confusion and indecision as to how to proceed in a practical and economical way in order to answer the numerous questions that still remain at the time of entry into the market.

It is important to realise at this point that the question is not necessarily global or universal, and that it has important connotations in regard to specific countries. This is due to the fact that some countries are traditionally allowing drugs into the market sooner than others (3). This analysis will therefore be done from a Canadian perspective, which takes into account the fact that most drugs have been marketed in the United States and/or the European Union from 6 to 12 months before being allowed access to the Canadian market (4-6). Of more than a theoretical interest is the question of whether Canadian patients benefit or suffer from this time lag.

II. WHAT DO WE KNOW OF THE DRUG ON THE DAY IT ENTERS THE CANADIAN MARKET?

Essentially, we have two sources of information. The first being the results of the Phase III trials and the second, the "real life" experience in the countries where the drug was marketed for a significant period of time before Canada. The Phase III trials are designed to demonstrate efficacy and to reassure regulators that the drug doesn't produce clinically severe and relatively frequent adverse events after short periods of exposure. They are done in very artificial settings on very carefully selected patients by very experienced teams of nurses and doctors. As a result of these conditions, they have very high internal validity but very low generalisability; their main drawbacks being the relatively small number of patients (3,000 to 5,000) and their short duration of treatment with the drug in question (6 to 12 months). They are therefore incapable of providing us with critical information on potentially severe side-effects (that occur on less than 1 out of 1,000 patients) that manifest themselves after more than one year of exposure. From a pharmaco-economic point of view, Phase III clinical trials also fail to provide information that is instrumental to the decisions that drug plan managers have to make. Placebo is commonly used as a comparator, whereas what is needed is a comparison with the cheapest or the most frequently prescribed drug for the same indication. In addition, and almost by definition, they provide no information on the most critical question drug plan managers have to answer: How is it going to be used? That is, for which indications other than the officially approved ones, at what doses, in which populations, and most important of all: which drug(s) is it going to replace?

The other source of information usually available in Canada at the time the drug enters the market is the experience gathered from countries where the drug has already been marketed. In terms of safety, several European countries and the United States have reasonably good and spontaneous reporting systems (5,7), which in the recent past have resulted in drugs being withdrawn from these markets before getting Canadian approval. Although it is reasonable to assume that a drug that produces undesirable side-effects in Europeans and/or Americans would also harm Canadian patients, it is much more difficult to extrapolate the drug use data obtained in countries foreign to Canadian settings. The prescription patterns and the availability of competing drugs is usually so different from one country to another that is it extremely risky to predict the Canadian drug use based on data obtained in other countries.
The concept of conditional approval of new drugs, which was discussed by Rawson, West and Appel (2) in a recent paper in the Journal of Clinical Pharmacology, is central to the eventual implementation of the post-marketing study of new drugs. The concept of conditional approval would provide a framework and a structure under which drugs will be studied after their entry into the market. From a Health Canada point of view, these studies would permit a better understanding of the safety of the drug and its long-term efficacy. From the point of view of the drug plan managers these studies would be invaluable in providing information that would allow them to verify whether the drug is used appropriately.

III. TYPES OF PHASE IV STUDIES

Phase IV studies can be divided in four classes: 1. the active pharmaco-vigilance cohort; 2. the prospective efficacy cohort; 3. the simplified clinical trial and; 4. the drug use study.

III.1 The active pharmaco-vigilance cohort

These are prospective patient cohorts who under their most simplified form can be considered as nothing else but patient registers. They allow a large number of patients to be followed for prolonged periods of time. Patients numbering up to 10,000 and treated for five years or more might be necessary to answer some questions, particularly those relating to safety. The object is to collect very specific information with a maximum degree of efficiency. From a practical point of view, this means a maximum of 20 questions with visits being no more frequent than every three months. These studies can be designed to provide additional information on specifically expected problems; for instance on drugs who in Phase III trial would have shown a low incidence of potentially serious problems, like the elevation of liver enzymes, allergic reactions or prolongation of the QT interval, which could be studied in such a cohort and thus provide information on whether these "red flags" will turn out to be predictors of rare but serious adverse events like hepatic necrosis, Stevens Johnson syndrome or sudden death. These prospective pharmaco-vigilance cohorts have to be flexible enough to be able to capture problems that were not suspected from the results of Phase III trials but could be fatal or life-threatening. Ideally, the collection and analysis of the data should be done in "real time" in order to allow a rapid response in cases where serious problems are encountered. These prospective pharmaco-vigilance cohorts have the great advantage of providing solid numerators and denominators, parameters that are notoriously fuzzy when provided by spontaneous case reports. A very practical problem relating to the implementation of an active pharmaco-vigilance cohort is to be able to distinguish them from "seeding" studies, which have contributed to give all Phase IV studies an undeserved bad reputation. The distinction is not that difficult. The implementation of a Phase IV study is legitimate if it answers an important public health question with an appropriate methodology.

III.2. The prospective efficacy cohort

The objective here is to demonstrate long-term efficacy. This is particularly important regarding drugs for which 6 to 12 months trials have demonstrated some degree of efficacy in an indication for which long-term effectiveness is notoriously difficult to obtain. Classical examples would be drugs used to treat obesity or to help in smoking cessation. The logistics can be extremely simple since the purpose would be to define whether the decreases in weight or if the rates of smoking cessation observed at the end of a six month trial are still present after two or three years.

III.3. The simplified clinical trial

This type of trial could be defined as a randomised controlled trial with a maximum degree of freedom as to how the patients are treated once randomisation has been implemented. These are extremely useful when the effectiveness (as opposed to the efficacy) of two drugs needs to be studied. Randomisation
becomes necessary when confounding by indication would create two groups whose clinical characteristics would be so different that the results could not be interpreted without complex, and sometimes not very credible, statistical manipulations. It is important that both patients and physicians are aware of which drug they have been assigned to. It is also critical that patients and their physicians have a maximum degree of leeway as to the frequency of the visits, the collection of laboratory tests and the treatment of adverse reactions from the drugs and/or complications of the disease. One of the main challenges of these studies is to convince the sponsors and the investigators, which have been brought up in the very structured classical Phase III trial, that useful information can be gathered without specific criteria about the frequency of the visits or the way blood pressure should be measured.

III.4. The drug use study

A critical question that drug plan managers would like to have answered at the time they decide on the inclusion of a new drug in their formulary is the following:

a. What is it going to replace?
b. A more specific form of this question would be: is the new drug going to be used as a first line or as a second line treatment?

These two questions are obviously not answered from the data available to the pharmaceutical company at the time the drug enters the market. Very hypothetical and speculative answers can derive from the knowledge of the medical practice, the companies marketing records and the way new drugs of the same class were previously used. The drug plan manager can only speculate on how the new drug will eventually be used. It is therefore perfectly reasonable to require drug use studies as a condition for listing new drugs, including the implicit acceptance by the pharmaceutical company that corrective measures will have to be taken in case the use of the drug happens to deviate excessively from what was predicted to be ideal in cost-efficient therapeutics. In addition, drug use studies can provide very useful information on the doses actually used in medical practice, which on some occasions can be much larger than those used in the clinical trials, and thus invalidate the initial pharmaco-economics studies based on the results of these trials. The appearance of the use of the drug for new, official or unofficial, indications can also dramatically invalidate the predictions of the pharmaco-economics studies based on Phase III data.

When the required information is contained in databases, these should be favoured since studies on databases have the advantage of being relatively cheap and fast. The other advantage of databases is that they permit one to obtain information on a population of doctors and patients who do not know they are being observed, and who therefore operate under their usual behavioural patterns. When the necessary information is not available in databases, it becomes necessary to implement field studies, which have the inconvenience of being much more expensive, but the enormous advantage of being designed specifically to capture all the necessary information. One of the problems with the prospective drug use studies done in the field is that the health professionals and the patients who agree to participate in the study might not be representative of the overall population. Another one is the fact that their knowing they are being observed might cause them to modify the way they practice. These two problems becoming extreme could render useless the interpretation of these studies because they would not represent the "real life" use of the drug under study.

VI. CONCLUSION

In conclusion, given their relative novelty, Phase IV studies constitute an absolute necessity for the protection of the patients and the proper use of the public funds used to reimburse drugs.
V. REFERENCES


Chapter 8. Bioavailability and Bioequivalence

Jake J. Thiessen, Ph.D.

Leslie Dan Faculty of Pharmacy
University of Toronto
Toronto, Ontario M5S 2S2
CANADA
Email jj.thiessen@utoronto.ca
I. INTRODUCTORY REMARKS

Over the last 25 years, Pharmacokinetics has emerged as an integral part of drug development, especially when identifying a drug's biological properties. By pharmacokinetics, one means the application of kinetics to a Pharmakon, the Greek word used to specify drugs and poisons. The term thereby implies the time course and fate of drugs in the body. This general definition broadly embraces absorption, distribution, metabolism (biotransformation) and excretion (ADME). The linking of Pharmacodynamics (response) and pharmacokinetics offers a composite understanding both about how the drug affects the body and how the body affects the drug.

The most comprehensive insight about a drug's inherent pharmacokinetic properties is gained by studying an intravenous dose. This route of administration has the greatest quantitative potential, as it permits a mass balance approach to be applied to distribution, clearance and the body processes associated with excretion and metabolic elimination (e.g. renal, hepatic). The administration of a drug by other routes, notably oral, introduces an uncertainty that reflects the unknown fraction that is actually absorbed. Consequently, such doses alone cannot accurately identify the distribution and clearance processes.

The most important property of any non-intravenous dosage form, intended to treat a systemic condition, is the ability to deliver the active ingredient to the bloodstream in an amount sufficient to cause the desired response. This property of a dosage form has historically been identified as physiologic availability, biologic availability or bioavailability. Bioavailability captures two essential features, namely how fast the drug enters the systemic circulation (rate of absorption) and how much of the nominal strength enters the body (extent of absorption). Given that the therapeutic effect is a function of the drug concentration in a patient's blood, these two properties of non-intravenous dosage forms are, in principle, important in identifying the response to a drug dose. Onset of response is linked to the rate of drug absorption whereas the time-dependent extent of response is linked to the extent of drug absorption. While the bioavailability of each type of non-intravenous product (e.g. oral, inhalation, topical (e.g. patch), rectal, etc.) could be discussed, this chapter will of necessity focus only on orally administered products. They certainly represent the major pharmaceutical class in drug development and patient treatment.

Bioavailability following oral doses may vary because of either patient-related or dosage-form-related factors. Patient factors can include the nature and timing of meals, age, disease, genetic traits and gastrointestinal physiology. The dosage form factors include 1) the chemical form of the drug (e.g. salt vs. acid), 2) its physical properties (e.g. crystal structure, particle size), and 3) an array of formulation (e.g. non-active ingredients) and manufacturing (e.g. tablet hardness) variables.

Not surprisingly, bioavailability is of clinical, academic, and regulatory interest. The latter includes agencies that approve the sale of products in their nation(s), as well as reimbursement agencies. Applications from manufacturers seeking regulatory approval for a new drug (e.g. New Drug Application (NDA)) must furnish exhaustive information about a drug's pharmacokinetics. Typically, such evidence entails studies wherein the drug has been orally administered. While such trials may broadly be viewed as bioavailability studies, many are ostensibly designed to assess the drug's safety and efficacy via strategies of dose escalation and chronic administration. These studies will not be entertained in this chapter. The more pertinent interest in bioavailability relates to questions about absolute extent of absorption (absolute bioavailability), the importance of product formulation changes that are made during a new drug's development process, the comparability of different oral dosage forms (e.g. modified-release versus conventional products), and whether the products can be administered with meals. These facets will receive attention in this chapter.

Manufacturers seeking regulatory approval of competitive (generic) products (e.g. Abbreviated New Drug Application [ANDA]), must provide detailed bioavailability evidence showing head-to-head comparative
II. COMPARATIVE BIOAVAILABILITY: A UNIVERSAL APPROACH

Most bioavailability studies, whether for a new or generic product, possess a common theme. A test is conducted to identify the quantitative nature of a specific product comparison. This comparison for a new drug may be, for example, to assess the performance of an oral formulation relative to that of an intravenous dose, or perhaps the performance of a modified-release formulation in comparison to a conventional capsule. For a generic product, it is typically a comparison of a competitive formulation with a reference product. Such commonality surrounding comparative bioavailability studies suggests a universal experimental approach.

All the studies to be described in this chapter basically attempt to establish a drug's concentration versus time profile following product administration in some form of comparative test. As shown in Figure 1, the two primary metrics for such concentration versus time profiles are the area under the curve (AUC) and the maximum observed concentration (Cmax); the former customarily includes the AUC to the last sampling time. The area to time infinity (AUC_{∞}) is the extrapolated area based on the AUC_{t} and the terminal constant (λ_{z}).
sampling time in a trial (AUC_t) and the extrapolated total AUC to time infinity (AUC_∞). The time at the maximum concentration (Tmax) is also of some minor interest.

After obtaining the profiles in a comparative trial, and computing the metrics, conclusions need to be reached regarding the comparison. Statistical methods are applied to test if the metrics are sufficiently similar to be considered equivalent. When the metrics are deemed equivalent, the drug concentration profiles are regarded as fundamentally the same. To achieve this equivalence, the study products' geometric mean ratios (eg. AUC_test/AUC_reference) and their computed 90% confidence intervals must reside completely within the 0.8 to 1.25 interval (See Figure 2).

**Figure 2:** An illustration of the statistical criteria to be satisfied to gain equivalence status in a comparative bioavailability assessment. For example, in a bioequivalence trial, the geometric mean ratio for the test/reference Cmax (GMR Cmax) must be located between 0.8 and 1.25. The GMR AUC's (whether AUC_t or AUC_∞) and their 90% confidence intervals must also be within the 0.8 to 1.25 window.

The preceding universal approach will be recognized as a common thread in the trials to be identified in this chapter. The data requirements for such an approach fundamentally orchestrate the design of the studies, which will be seen to have a rather common or universal format.

**III. COMPARATIVE BIOAVAILABILITY STUDIES FOR NEW DRUGS (NDA)**

The initial oral formulation for a new drug is frequently used to conduct early human studies of safety and efficacy. Often, early oral bioavailability information about the drug (and this initial formulation) is
obtained by means of studies comparing it with an intravenous dose and/or a solution of the drug. Although such studies will not be described in this chapter, they employ the Universal Approach wherein the comparator is an intravenous dose or perhaps a solution of the drug.

When the oral dosage formulation undergoes changes during the drug development process, the deductive inference concept becomes a helpful tool. It circumvents the need to retest subjects or patients following each formulation change in order to reestablish product safety and efficacy. The fundamental tenet underpinning the logic is similar to that described later for generic product testing. First, it is assumed that the time-dependent drug concentrations in blood from an early formulation are intimately linked with the effects. Second, if a new formulation exhibits the same time-dependent drug concentrations (rate and extent of drug absorption), the new formulation is deemed "bioequivalent" and, by inference, has the same safety and efficacy.

To test reformulated dosage forms, the Universal Approach is employed. The fundamental nature of the study is similar to that described in detail within the "Bioequivalence" section of this chapter.

III.1. Outline of a typical product reformulation bioavailability study:

III.1.A. Objectives
To test the comparative bioavailability of a reformulated and original product and thereby to determine their equivalence.

III.1.B. Primary endpoints
Determine the time-dependent concentrations of the administered drug in the collected blood (or plasma/serum) of each subject following administration of the reformulated and original products.

III.1.C. Secondary endpoints
Determine the time-dependent concentrations of potentially important metabolites (active and contributing to the product's therapeutic response) in the collected blood (or plasma/serum) of each subject following administration of the reformulated and original products.

III.1.D. Exploratory endpoints
Determine the Cmax, AUC∞, Tmax, λz and half-life of the primary (and secondary) endpoints following the reformulated and original products, for each subject.

III.1.E. Study design
The study shall be designed in such a way that the effects of formulation can be distinguished from other factors. When two formulations are compared, a randomized two-period, two-sequence crossover study is considered the design of choice. An adequate washout period between periods is needed to avoid drug carryover effects.

All facets of the study are to be tightly controlled. Normally, subjects fast for 10 hours prior to product administration. A dose of the tested product is administered at the start of an experimental day with about 8 ounces (240 mL) of water. Further fluid will be withheld for about 2 hours; standardized meals are permitted beginning at four hours after drug administration. All subsequent meals will be carefully standardized at fixed intervals.

Sequential blood samples (about 12 to 18, including a pre-dose sample) shall be drawn at appropriate, specified, and carefully recorded times (to capture increasing and decreasing concentrations during the absorption, distribution and elimination phases). The collections are to continue for about three terminal drug half-lives in order to capture at least 80% of the total area. At least three to four samples need to be
obtained from the terminal log-linear phase to derive an acceptable estimate of the terminal constant ($\lambda_z$) from linear regression. For long half-life drugs, a truncated AUC (e.g. up to 72 hours) is generally considered adequate.

Blood samples or the harvested plasma/serum shall be analyzed for the administered drug or metabolites by means of a validated analytical method.

III.1.F. Planned sample
The appropriate subject number can be forecast, as described in the "Bioequivalence" section, via the ANOVA error variance associated with the specific metric (e.g. from an earlier study), the expected deviation of the reformulated products' metrics and the bioequivalence criterion (e.g. 90% confidence that the estimated population mean ratio lies between 80 and 125%).

III.1.G. Study population
Healthy volunteers are normally selected, although for some drugs it may, of necessity, be best to conduct the trial in patients. See the "Bioequivalence" section.

III.1.H. Specific inclusion criteria
Healthy males or females will be included in the study population. Preferably, non-smokers will be employed.

III.1.I. Specific exclusion criteria
Women of childbearing potential are to be excluded if there is a potential risk. Other common exclusion criteria are identified in the "Bioequivalence" section.

III.1.J. Tools for assessing primary endpoints
A validated analytical method is needed for both the primary and secondary endpoints.

III.1.K. Specific criteria for early withdrawal and discontinuation
While the number and availability of subjects shall be sufficient to allow all periods of the study to be successfully completed without coercion, subjects shall retain the right to discontinue the trial. Discontinuation reasons may include adverse drug reactions or even personal preferences. All withdrawals must be reported.

III.1.L. Data analysis method:
Consult the Universal Approach (Figure 2) and the "Bioequivalence" section. In summary, ANOVA is to be used to identify the source contributions by factors including subjects, period, formulation and potential interactions. The geometric mean ratio together with the ANOVA residual mean error term are used to identify the statistical basis for the 90% confidence interval for the ratio of the population means (New Formulation/Original Formulation) of the identified metrics (e.g. AUC, Cmax).

III.2. Development of a new formulation (e.g. modified release product)
Delayed-release products typically release the active ingredient at a time later than immediately after administration, thereby sometimes exhibiting an absorption lag time. The first modified-release product requires an NDA. The purpose of the required studies is to determine if the following conditions are met:
   a. The drug product meets the controlled release claims made for it;
   b. The bioavailability profile rules out the occurrence of what is called "dose dumping", which is the premature release of the drug from the dosage form;
   c. The formulation provides consistent performance between individual dosage units;

60
d. The steady state performance, in comparison to an available conventional product, is equivalent. If, based on accumulated evidence between circulating concentrations of the drug and response, the modified-release product is different, clinical studies will be needed to show the impact of such differences.

The study requirements for modified-release products permit some flexibility, but shall include the following:

a. A single dose crossover comparison of a conventional, immediate release, product and the modified release product (ideally, the study would also include a solution or suspension of the same drug in the same strength);
b. A single dose food-effect study;
c. A steady-state study.

For the first two study requirements above, the Universal Approach is again needed. While summary information about the food effect study is presented here, further details of a food effect study are presented in the "Bioequivalence" section. The steady-state study requirements will not be presented in this chapter because they do not have a similar "comparative" character. The primary requirements for the two comparative studies are:

III.3. Objective A

To test the comparative bioavailability of a modified-release and immediate-release product and thereby to determine their equivalence.

III.3.A. Primary endpoints
Determine the time-dependent concentrations of the administered drug in the collected blood (or plasma/serum) of each subject following administration of the modified-release and immediate-release products.

III.3.B. Exploratory endpoints
Determine the Cmax, AUCt, AUC∞, Tmax, λz and half-life of the primary (and secondary) endpoints following the modified-release and immediate-release products, for each subject. Some agencies will also require the area over the usual dosing interval for the modified-release product.

III.3.C. Study elements
The fasting study shall be designed in such a way that the effects of the formulation can be distinguished from other factors. When two formulations are compared, a randomized two-period, two-sequence crossover study is considered the design of choice. An adequate washout period between periods is needed to avoid drug carryover effects.

Doses are given to subjects following an overnight fast. All remaining aspects are to be controlled as outlined in the "Bioequivalence" section. Data analysis is also conducted as described in that section. However, it should be recognized that differences in Cmax can be anticipated because the fundamental drug release properties for the modified-release and immediate-release products are different. The potential impact of such differences needs to be weighed in the light of concentration versus response evidence.

III.4 Objective B

To test the effect of food upon the bioavailability of a modified-release product.
III.4.A. Primary endpoints
Determine the time-dependent concentrations of the administered drug in the collected blood (or plasma/serum) of each subject following administration of the product under fasting and fed conditions.

III.4.B. Exploratory endpoints
Determine the Cmax, AUCt, AUC∞, Tmax, λz and half-life of the primary (and secondary) endpoints following administration of the modified-release product in each period, for each subject.

III.4.C. Study elements
This fasting/fed study shall be designed in such a way that the potential effects of the meal upon the formulation can be distinguished from other factors. When the formulation is tested as required, a randomized two-period, two-sequence, crossover study is considered the design of choice. An adequate washout period between periods is needed to avoid drug carryover effects.

Doses are given to subjects following an overnight fast. In one period, the fast is continued, whereas in the other period a meal is given before dose administration. For meal administration, please refer to the food-effect study described in the "Bioequivalence" section. Typically, both fasting and fed periods become common four hours after dose administration when normal food ingestion cycles are permitted.

All study aspects are controlled as outlined in the "Bioequivalence" section. Data analysis is also conducted as described in that section and under "Universal Approach".

IV. COMPARATIVE BIOAVAILABILITY FOR GENERIC DRUG PRODUCTS (ANDA): BIOEQUIVALENCE STUDIES

The deductive inference concept is also central to bioequivalence testing. The foundation is set, first, through evidence that a specified, approved, reference drug product (e.g. tablet from the innovative manufacturer) has shown acceptable safety and efficacy through an array of clinical trials. Second, a widely held view is embraced that the time-dependent drug concentrations in blood from such a reference product are intimately linked with the therapeutic effects. Third, a principle is adopted, namely that chemically equivalent (same amount of the same active ingredient) and pharmaceutically equivalent products (same dosage form; e.g. conventional tablet), that exhibit the same rate and extent of drug absorption, are bioequivalent. Fourth, bioequivalent products by inference are considered therapeutically equivalent.

When a manufacturer thereby wishes to gain therapeutic equivalence by introducing a competitive generic product into the marketplace, it is not necessary to conduct the full array of trials needed for the first (innovative) product. If equivalence has been demonstrated, according to prescribed study requirements, appropriately determined metrics (Figure 1), and statistical criteria (Figure 2), the generic product by inference is regarded as therapeutically equivalent to the innovative drug product.

The design of and requirements in, bioequivalence studies are fundamentally satisfied through single dose administrations, although there is a lingering interest in multiple dose testing. The focus is on the rate and extent of absorption of the active ingredient, although some jurisdictions (e.g. FDA) continue to show an interest in the primary active metabolite(s). In some cases, notably drugs that exhibit non-linear pharmacokinetics, the dose strength to be tested may be dictated by whether the drug's non-linearity is attributable to the absorption or elimination phase (Health Canada). As a general principle, the studies are designed to test inherent product absorption properties. Thereby, the trials generally specify healthy normal controls that exhibit circumscribed demographics.
Comparative evidence may require not only studies in a fasting condition, but following a specified meal. The latter permit drug formulations to be evaluated under "stressed conditions". If it is shown that competitive products are bioequivalent under both fasting and fed conditions, there is greater confidence that they are therapeutically equivalent when used in patients.

IV.1 Testing competitive (generic) products under fasting conditions

The following describes the requirements for most orally administered products, including tablets, capsules and modified-release dosage forms. Nevertheless, it is best to check with each regulatory agency regarding current or special drug- or product-specific requirements.

The bioequivalence study conclusions are commonly extended to all strengths of the products provided the active and inactive ingredients conform to regulatory requirements of proportionality. When these requirements are violated, representative strengths of each formulation type shall be tested.

IV.1.A. Objectives
To test the comparative bioavailability of a test and reference product and thereby to determine their equivalence.

IV.1.B. Primary endpoints
Determine the time-dependent concentrations of the administered drug in the collected blood (or plasma/serum) of each subject following administration of the test and reference products.

IV.1.C. Secondary endpoints
Determine the time-dependent concentrations of potential important metabolites (active and contributing to the product's therapeutic response) in the collected blood (or plasma/serum) of each subject following administration of the test and reference products.

IV.1.D. Exploratory endpoints:
Determine the Cmax, AUCt, AUC∞, Tmax, λz and half-life of the primary (and secondary) endpoints following each of the test and reference products, for each subject.

IV.1.E. Study design
The study shall be designed in such a way that the effects of formulation can be distinguished from other factors. If two formulations are compared, a randomized two-period, two-sequence crossover study is considered the design of choice. An adequate washout period between periods is needed to avoid drug carryover effects. Replicate studies, although not mandated, offer the advantage of providing a comparison of intra-subject variances for the test and reference products.

All facets of the study are to be tightly controlled. The full characteristics, including lot numbers and expiry dates of the test and reference products, shall be known. Normally, subjects fast for 10 hours prior to product administration. Normally, the highest safe strength/dose of the test or reference product will be administered at the start of an experimental day with about 8 ounces (240 mL) of water. Further fluid will be withheld for about 2 hours; standardized meals are to be permitted beginning at four hours after drug administration. All subsequent meals will be carefully standardized according to a fixed schedule.

For most drugs, subjects shall not be permitted to recline until at least two hours after product ingestion. Physical activity and posture is to be standardized to limit variable effects on gastrointestinal blood flow and motility. Blood samples (about 12 to 18, including a pre-dose sample) shall be drawn at appropriate, specified, and carefully recorded times (to capture increasing and decreasing concentrations during the absorption, distribution and elimination phases). The collections are to continue for about three terminal
drug half-lives in order to capture at least 80% of the total area. At least three to four samples need to be obtained from the terminal log-linear phase to derive an acceptable estimate of the terminal constant \((\lambda_z)\) from linear regression. For long half-life drugs, a truncated AUC (e.g. up to 72 hours) is generally considered adequate.

Blood samples or the harvested plasma/serum shall be analyzed for the administered drug or metabolites by means of a validated analytical method.

**IV.1.F. Planned sample**
While most jurisdictions support a minimum of 12 subjects in a bioequivalence trial, the likelihood of a successful outcome is improved with an increase in the subject number. The appropriate subject number can be forecast via the ANOVA error variance associated with the specific metric (e.g. from published data or a pilot study), the expected deviation of the test product's metric from that of the reference product (e.g. 0.05) and the bioequivalence criterion (e.g. 90% confidence that the estimated population mean ratio lies between 80 and 125%).

**IV.1.G. Study population**
To minimize variability and focus on the comparison of the two formulations, healthy volunteers are to be selected, although for some drugs it may, of necessity, be best to conduct the trial in patients. Subjects will ordinarily be between 18 and 55 years of age and within the accepted normal range for Body Mass Index. Clinical laboratory tests, notably to assess cardiac, renal and hepatic function, are to be normal based on subject screening. Furthermore, subjects will have undergone an extensive review of medical history and received a comprehensive medical examination.

**IV.1.H. Specific inclusion criteria**
Healthy males or females will be included in the study population. Preferably, non-smokers will be employed.

**IV.1.I. Specific exclusion criteria**
Women of childbearing potential are to be excluded if there is a potential risk. Subjects shall not have a history of alcohol or drug abuse. Subjects shall not be receiving drugs for any medical condition. There is to be no known allergy to the administered drug or formulation. As a rule, alcoholic beverages and over-the-counter drugs shall be avoided during the days immediately preceding a trial and for an appropriate interval during the active sample collection period of the trial.

**IV.1.J. Tools for assessing primary endpoints**
A validated analytical method is needed for both the primary and secondary endpoints.

**IV.1.K. Specific criteria for early withdrawal and discontinuation**
While the number and availability of subjects shall be sufficient to allow all periods of the study to be successfully completed without coercion, subjects shall retain the right to discontinue the trial. Discontinuation reasons may include adverse drug reactions or even personal preferences. All withdrawals must be reported.

**IV.1.L. Data analysis method**
All study information, including exploratory endpoints shall be presented for each subject following the test and reference products. ANOVA is to be used to identify the source contributions by factors including subjects, period, formulation and potential interactions. The geometric mean ratio together with the ANOVA residual mean error term are used to identify the statistical basis for the 90% confidence interval for the ratio of the population means (Test/Reference) of the identified metrics (e.g. AUC, Cmax).
Health Canada's Part A Guide (See "Suggested Readings") provides an amplified section illustrating the calculations.

**IV.2. Testing competitive (generic) products under fed conditions**

Food-effect studies are recommended particularly for modified-release dosage forms and, in some jurisdictions, for an array of conventional solid oral products.

Commonly, aside from the incorporation of a meal, the same testing methods are to be used as described above for the fasting condition. Therefore, only the study design is presented below.

**IV.2.A. Study design**

The fed study is to be designed in such a way that the effects of formulation can be distinguished from other factors. If two formulations are being compared, a randomized two-period, two-sequence crossover study is commonly considered the design of choice. An adequate washout period between periods is needed to avoid drug carryover effects. Replicate studies, although not mandated, offer the advantage of providing a comparison of intra-subject variances for the test and reference products.

All facets of the study are to be tightly controlled. The full characteristics, including lot numbers and expiry dates, of the test and reference products shall be known. Normally, subjects fast for 10 hours prior to ingesting a standardized meal. The meal is to provide the greatest changes from the gastrointestinal physiology of a fasting state. A meal with high-fat and high-calorie content is recommended (e.g. 150, 250 and 500-600 calories from protein, carbohydrate, and fat, respectively). The meal shall be ingested over a period of 30 minutes or less. The product dose shall be ingested 30 minutes after start of the meal.

Generally, the highest safe strength/dose of the test or reference product will be administered with about 8 ounces (240 mL) of water. Further fluid shall be withheld for about 2 hours; standardized meals will be permitted beginning at four hours after drug administration. All subsequent meals will be carefully standardized.

For most drugs, subjects shall not be allowed to recline until at least two hours after product ingestion. Physical activity and posture shall be standardized to limit effects on gastrointestinal blood flow and motility. Blood samples (about 12 to 18, including a pre-dose sample) are to be drawn at appropriate, specified, and carefully recorded times (to capture increasing and decreasing concentrations during the absorption, distribution and elimination phases). The collections shall continue for about three terminal drug half-lives in order to capture at least 80% of the total area. At least three to four samples shall be obtained from the terminal log-linear phase to derive an acceptable estimate of the terminal constant ($\lambda_z$) from linear regression. For long half-life drugs, a truncated AUC (e.g. up to 72 hours) is generally considered adequate.

Blood samples or the harvested plasma/serum are to be analyzed for the administered drug or metabolites by means of a validated analytical method.

**V. REPRESENTATIVE WELL DESIGNED TRIALS**


VI. SUGGESTED READINGS

5. EMEA (European Agency for the Evaluation of Medicinal Products), CPMP (Committee for Proprietary Medicinal Products). Note for guidance on the investigation of bioavailability and bioequivalence. 2001.
Chapter 9. Pharmacoeconomics and Economic Evaluation of Drug Therapies

Tom Walley, M.D.

Professor of Clinical Pharmacology
Department of Pharmacology and Therapeutics
University of Liverpool
70 Pembrooke Place
Liverpool L69 3BX
UK
Tel 0151 794 8123
Email twalley@liv.ac.uk
I. INTRODUCTORY REMARKS

Health care funders (governments, social security funds, insurance companies) are struggling to meet their rising costs. They make many efforts to contain drug costs, by price negotiation, patient co-payments or dedicated drug budgets. Expenditure on drug therapy is a particular target for their attention for several reasons: the size of the drug bill (10-15% of most national health care budgets, and usually the second largest item after salaries); the ease of measurement of pharmaceutical costs in isolation, in contrast to most other health care costs; evidence of wasteful prescribing; and a perception that many drugs are overpriced and that the profits of the pharmaceutical industry are excessive.

But this focuses on drug costs in isolation, when what should be of greater concern to decision makers, health care professionals and the public is the value of drug therapy, a function of its benefits as well as its costs (1). Payers acknowledge that spending on drugs, which may for instance reduce the need for hospitalisation or produce greater health gain for the same resources than other medical interventions, may be a very efficient use of scarce resources. They therefore respond by demanding evidence of value for money from drug therapy. Drug therapy is open to this simply because there are high quality trials to support most new drugs’ licensing applications, in contrast to the poor evidence around most other health care interventions.

Health economics is the science of assessing cost and benefits, not to make decisions about resource use, but to inform those decisions. The aim is to identify what is most efficient, so that the greatest amount of benefit can be bought for a given amount of money or resources. But we must remember that in health care, efficiency may not be the most important objective—we might for instance prioritise caring for dying patients or treating patients with serious disease who have relatively little hope of surviving. Pharmacoeconomics is a branch of health economics that particularly considers drug therapy. It is of particular interest to pharmaceutical companies who in developing a new drug and after the traditional hurdles of efficacy, safety and tolerability must now jump over a fourth hurdle of cost effectiveness. It should also interest clinical pharmacologists, either in their roles assessing new drugs or in the conduct of clinical trials that now often include an economic component. In some areas, economic studies have become an accepted part of evaluations for reimbursement.

This chapter aims to explain the basic concepts and language of pharmacoeconomics, and of economic evaluation, and to introduce what for many clinical pharmacologists is a new area.

II. BASIC CONCEPTS AND TERMINOLOGY

Health economics is about making choices between options, when there is scarcity of resources. It is fundamentally comparative, weighing the costs and benefits of option 1 with those of option 2 (for instance, a new drug and the previous best therapy - traditional medical evaluation focused only on the benefits), to determine which is the most efficient way to use our limited resources. Efficiency is a key concept in economics, i.e. how to buy the greatest amount of benefit for a given resource use.

II.1. Opportunity cost

Another key concept is opportunity cost: this is defined as the “benefit foregone when selecting one therapy alternative over the next best alternative”. When we have limited money and we spend it on one health care intervention, we cannot spend the same money on something else. So we should be less concerned with how much a health care intervention costs, but rather with what other benefits we are giving up by using the money in that way. We need to be sure that spending money on the new therapy will buy more benefit than spending that money in some other part of the health care system.
The comparative nature of health economics means that we are interested in an Incremental analysis of costs and benefits. There is usually a current treatment for most conditions, with associated costs and benefits. We would not advocate stopping all existing treatment for the condition, so the question is not what are the costs and benefits of the new treatment, but what are its added costs and benefits, over and above those of the existing treatment.

II.2. Marginal cost

A related concept is marginal costs. For instance, if a new treatment enables patients to be discharged from hospital a day earlier than an older treatment, it might be tempting to count the average cost of a hospital bed day as a saving of resources. But all the fixed capital charges for a hospital bed, which go into the average cost, e.g. costs of laboratories, kitchens, and building maintenance, will be largely unchanged. The only costs which change may be those of having a patient physically occupy the bed - the costs of the patient’s meals, treatment and perhaps nursing time. These are the marginal costs, where the resource use actually changes substantially. Incremental analysis is concerned with the marginal and not the average costs. Marginal costs are often very difficult to measure, and there is a temptation to use average costs instead. This may be justified if for instance, enough bed days are saved by the widespread adoption of a new treatment to actually reduce bed numbers and to close wards.

III. COSTS AND BENEFITS

These have broad definitions in health economics, which may depend in part on the perspective or viewpoint we choose to take. Perspective asks from whose point of view is the study conducted - from that of the health care payer, who is only interested in the direct costs of health care, or from society as a whole, where “indirect” costs (i.e. not directly on health care, such as lost of productivity etc) are also important. In general, the societal perspective is considered the most appropriate, but a health care manager with a limited budget might be tempted to ignore the societal view and consider only the costs that fall on his own budget. A study of migraine which took the health service perspective only might suggest that sumatriptan in migraine (an expensive drug in an area which previously cost the health service very little) was highly undesirable, but a study taking a societal perspective might come to the opposite conclusion (2).

III.1. Cost classification

Costs therefore can be classified as:

III.1.A. Direct – i.e. costs from the perspective of the healthcare funder: including staff costs, capital costs, drug acquisition costs. These should (in theory) be relatively easy to measure.

III.1.B. Indirect – i.e. costs from the perspective of society as a whole: for example, these might include loss of earnings, loss of productivity, loss of leisure time, due to the illness, and cost of travel to hospital etc. This would include not just the patient themselves but also their family and society as a whole. Many of these are difficult to measure, and there is some controversy over how to value these. (The UK National Institute for Clinical Excellence, NICE, adopts a limited societal perspective in its evaluations and considers the direct costs falling on the UK National Health Services, and those indirect costs funded by the state such as unemployment and sickness benefits (3)).

III.1.C. Intangible – i.e. the pain, worry or other distress which a patient or their family might suffer. These may be impossible to measure in monetary terms, but are sometimes captured in measures of quality of life.
III.2. Benefits

The benefits we expect from an intervention might be measured in:

III.2.A. “Natural” units - e.g. years of life saved, strokes prevented, peptic ulcers healed etc.

III.2.B. “Utility” units - utility is an economist’s word for satisfaction, or sense of well being, and is an attempt to evaluate the quality of a state of health, and not just its quantity. Utility estimates can be obtained through direct measurement (using techniques such as time trade off or standard gambles, or by imputing them from the literature or expert opinion. They are often informed by measures of quality of life in different disease states.

The Quality Adjusted Life Year (QALY) is one widely used measure, which attempts to integrate both quality and the quantity of life. Broadly, it assumes that if a treatment increases one’s life expectancy by 2 years, but causes adverse effects or inconvenience, such that one’s quality of life or utility are decreased by 25%, the net gain is $2 \times 0.75 = 1.5$ QALYs. QALYs are controversial for many reasons (4), not least that measuring patient utilities is difficult and preferences may change in the course of an illness (what seems an intolerable burden to a healthy individual may not seem so bad to someone who might otherwise be dead). Despite these criticisms, the concept of the QALY has advanced thinking on how to incorporate quality of life into economic evaluations.

### Table 1.

<table>
<thead>
<tr>
<th>Method of economic evaluation</th>
<th>Measurement of outcome (health benefits)</th>
<th>Synthesis of costs and benefits</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cost minimisation analysis</td>
<td>Assumed to be equivalent and can take any form (e.g. number of cases detected, reductions in cholesterol levels, years of life saved)</td>
<td>Additional costs of therapy A relative to B</td>
</tr>
<tr>
<td>Cost effectiveness analysis</td>
<td>Health benefits across therapies are measured in similar natural units</td>
<td>Cost per life year gained, Cost per patient cured, Cost per life saved, etc.</td>
</tr>
<tr>
<td>Cost utility analysis</td>
<td>Health benefits across therapies are valued in similar units based on individual preferences</td>
<td>Cost per QALY gained, Cost per HYE gained</td>
</tr>
<tr>
<td>Cost benefit analysis</td>
<td>Measured in similar or different units and are always valued in monetary units (e.g., amount willing to pay to prevent a death, amount willing to pay to reduce exposure to a hazard)</td>
<td>Net benefits = Benefits minus costs, Benefit-cost ratio = benefits/costs</td>
</tr>
</tbody>
</table>

III.2.C. Associated Economic Benefit

This is usually measured in money, which is a useful common denominator allowing comparisons across different disciplines. This measure includes, for instance, the economic benefits of returning someone to work.

III.2.D. Methods of Economic Evaluation

Economic evaluation is the formal process of weighing benefits and costs in an incremental analysis. It is essentially a framework which draws up a balance sheet between costs and benefits to assist decision making.

*Common Types of Study*

The costs and benefits or outcome measures selected give rise to the four common types of economic evaluation (table 1). These studies are often complex and require use of economic models (a skill not dissimilar to pharmacokinetic modelling).
III.2.D.a. Cost minimisation analysis (CMA)
This involves measuring only costs, usually only to the health service, and is applicable only where the outcomes are identical and need not be considered separately. An example would be prescribing a generic preparation instead of the brand leader (lower cost but same health outcomes).

III.2.D.b. Cost effectiveness analysis (CEA)
The term cost effectiveness is often used loosely to refer to the whole of economic evaluation, but should properly refer to a particular type of evaluation, in which the health benefit can be defined and measured in natural units (e.g., years of life saved, ulcers healed) and the costs are measured in money. It therefore compares therapies with qualitatively similar outcomes in a particular therapeutic area. For instance, in severe reflux oesophagitis, we could consider the costs per patient relieved of symptoms using a proton pump inhibitor compared to those using H₂ blockers. CEA is the most commonly applied form of economic analysis in the literature, and especially in drug therapy. It does not allow comparisons to be made between two totally different areas of medicine with different outcomes. The broad form of these evaluations are shown in box 1, and the key measure is the incremental cost effectiveness ratio (ICER).

Box 1.

<table>
<thead>
<tr>
<th>Incremental Cost Effectiveness Ratio =</th>
<th>(cost of drug A - cost of drug B)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(benefits of drug A - benefits of drug B)</td>
</tr>
<tr>
<td>ICER =</td>
<td>difference in costs (A-B)</td>
</tr>
<tr>
<td></td>
<td>difference in benefits (A-B)</td>
</tr>
</tbody>
</table>

III.2.D.c. Cost utility analysis (CUA):
This is similar to cost effectiveness in that the costs are measured in money and there is a defined outcome (box 2). But here the outcome is a unit of utility (e.g. a QALY). Since this endpoint is not directly dependent on the disease state, CUA can in theory look at more than one area of medicine, e.g., cost per QALY of coronary artery bypass grafting versus cost per QALY for erythropoietin in renal disease. In practice this is not so easy since the QALY is not a well-defined fixed unit transferable from study to study. We should be particularly wary of attempts to draw up league tables of QALYs to allow comparisons between a range of therapies. The values in such tables have usually been derived at different times and in different ways and are not comparable.

Box 2. Calculating QALYs - a simple example

<table>
<thead>
<tr>
<th>With treatment X</th>
<th>Without treatment X</th>
</tr>
</thead>
<tbody>
<tr>
<td>Estimated survival = 10 years</td>
<td>Estimated survival = 5 years</td>
</tr>
<tr>
<td>Estimated quality of life (relative to ‘perfect health’) = 0.7</td>
<td>Estimated quality of life (relative to ‘perfect health’) = 0.5</td>
</tr>
<tr>
<td>QALYs = (10 x 0.7) = 7.0</td>
<td>QALYs = (5 x 0.5) = 2.5</td>
</tr>
<tr>
<td>QALY gain from treatment X = 7 - 2.5 = 4.5 QALYs</td>
<td></td>
</tr>
</tbody>
</table>

If the cost of treatment X is £18,000 then the cost per QALY is £4,000 per QALY (£18,000 divided between 4.5 additional QALY’s)
III.2.D.d. Cost benefit analysis (CBA)
Here, the benefit is measured as the associated economic benefit of an intervention (e.g. monetary value of returning a worker to employment earlier), and hence both costs and benefits are expressed in money. CBA may ignore many intangible but very important benefits not measurable in money terms, e.g. relief of anxiety. CBA may also seem to discriminate against those in whom a return to productive employment is unlikely, e.g. the elderly, or the unemployed.

However the virtue of this analysis is that it may allow comparisons to be made between very different areas, and not just medical, e.g. cost benefits of expanding university education (benefits of improved education and hence productivity) compared to establishing a back pain service (enhancing productivity by returning patients to work). This approach is not widely used in health economics, although many economists like it on theoretical grounds and because it removes some of the “sacred cow” protection which surrounds health care. They argue that health should be another commodity, and not necessarily valued more than other possible uses of the resources.

III.2.D.e. Cost consequences and other types of evaluation
Other forms of quasi-health economic evaluation may be seen in the literature but are not true economic evaluations because they do not weigh costs and benefits in an incremental manner. In some cases, often where studies consider multiple outcomes, costs and benefits are presented in a disaggregated form (e.g. health profiles). These evaluations are frequently referred to as cost consequences analyses. Burden of disease (also known as cost of illness) studies attempt to measure the health and resource implications arising to society from a particular disease.

III.3. Further Points

There are two further points for definition.

III.3.A. Discounting
There is often a difference in timing between the investment of health resources and gaining the benefits. Therefore we must discount future spending etc. to try to equalise the effects of inflation and health and financial preferences over a long period. In general, costs are discounted at an agreed rate (in the UK, currently 6% for costs). There is some debate over whether benefits can also be discounted (it is relatively easy to accept that £100 spent now is worth more than in five years time, but how does one compare a healthy year now to a healthy year in five years time?) NICE suggests discounting benefits at a rate of 1.5% (3).

III.3.B. Handling uncertainty
The measures of benefit and cost in an economic evaluation come from the medical evidence, usually clinical trials. But clinical trials address efficacy whereas health economics is more interested in effectiveness – what benefits/costs are associated with a new therapy when it is used in the real world, where patients are less well defined or monitored and where the comparator may not be the one used in the clinical trial. There is often little evidence available about effectiveness, and we are forced to make assumptions to fill the gaps in our knowledge. These assumptions should be reasonable, and should be transparent, so that they can be challenged. Any good economic study will challenge these assumptions itself, by varying them in a sensitivity analysis. This explores the extent to which a conclusion is dependent on an assumption. For instance if a study assumes a rate of relapse of duodenal ulcers after treatment of 5% at one year, what happens if the relapse rate were to be actually 2.5%, or 10%? This might drastically affect the outcome of a study. A sensitivity analysis clarifies what are the critical assumptions and confirm that the results of the evaluation are robust, despite changes in the assumption.
IV. HANDLING THE RESULTS OF ECONOMIC EVALUATIONS

Consider the four possible results arising in a CEA (figure 1). First, if costs are lower and health benefits higher for one drug relative to another, the former is said to dominate and would be the preferred treatment (quadrant II). Second, the opposite applies, i.e. the new drug is more expensive and less effective, and thus is considered inferior and not recommended (quadrant IV). The third and most common case is where the new drug is both more effective and more expensive than the standard (quadrant I); on the basis of ICERs, a judgement must be made regarding whether the additional benefits are worth the extra costs of the new drug and, therefore, whether it is ‘cost-effective’. This might be defined by a previously agreed ICER threshold value. The fourth case is similar to the third, with the roles of the new therapy and the standard reversed (quadrant III); the question now is whether the extra benefits provided by the standard justify the additional costs of retaining it as the preferred treatment when the option of a new, cheaper but less effective drug exists.

Figure 1.

**Difference in costs**

<table>
<thead>
<tr>
<th>IV</th>
<th>I</th>
</tr>
</thead>
<tbody>
<tr>
<td>The new treatment is less effective and more expensive</td>
<td>The new treatment is more effective and more expensive</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>III</th>
<th>II</th>
</tr>
</thead>
<tbody>
<tr>
<td>The new treatment is less effective and less expensive</td>
<td>The new treatment is more effective and less expensive</td>
</tr>
</tbody>
</table>

Defining what is an acceptable maximum value or threshold for ICER is difficult and controversial, since it clearly carries an element of rationing of care. How much is an extra QALY or life year worth? This is a value judgment. It can be explored to some extent through techniques such as trying to identify what a patient or the public might be willing to pay to avoid an unfavourable outcome. There may be precedents – e.g. by common consent, we provide treatment in the form of coronary bypass grafting: we work out later that this cost £X per QALY, and so this establishes a baseline for our thinking about how much we value a QALY. In the UK, NICE seems to operate at a threshold of around £30,000 per QALY (5), although no formal threshold is declared and its existence has been formally denied. One might be more confident in setting a threshold if economic evaluations were more certain in their outcomes.

V. LIMITS OF PHARMACOECONOMIC EVALUATION

Many problems limit our use of health economics in practice (1). The whole process may be open to bias, in the choice of comparator drug, the assumptions made, or in the selective reporting of results. This suspicion arises because most studies are conducted or funded by pharmaceutical companies who
obviously are interested in the results, and there is a publication bias towards those studies favourable to
sponsoring companies (6). Health economics is therefore sometimes misused as a marketing ploy. The
same problems may however arise in studies funded by health care payers. To a specialist, this is not such
a problem since the almost inevitable biases are usually clear. But since economic evaluation is less well
understood by doctors and others, bias needs to be minimised.

Doctors may tend to equate health economics with rationing or cost cutting, and many therefore reject on
principle the whole process as unethical. Since resources are limited within health services, wasting them
by inefficiency is wrong, as it reduces the clinician’s ability to give the best possible care to his patients.
It therefore seems unethical not to consider the economics of a medical intervention.

A key problem is our ability to implement the results of a study. No matter how good a study is, and how
cost effective a therapy compared to existing treatment, it may not be possible to achieve its potential
benefits because of the existing cumbersome management structures. Three problems are common: first, a
short term outlook which limits the application of economic evaluations showing long term savings for the
health service in return for increased spending now. Second, many budgets operate in isolation, and it is
not easy to move money between them: for instance, prescribing in primary care is often funded separately
from hospital services, so any increased spending on drug therapy in primary care cannot be simply
funded from a future reduction in hospital admissions. Third, a new intervention may simply not be
affordable no matter how cost effective it might be.

Finally, health economics and pharmacoeconomics is a young science and is slowly developing and
testing its methodologies. We do not have space to address all of these concerns here but many of the
details of the methods described above are academically and practically controversial (2). There have
been many guidelines developed (e.g. ref 3) for the conduct of economic evaluation, recognising the
possibilities of bias and the poor understanding of many potential users about the whole process.

VI. THE FUTURE

Despite these problems, economic evaluations of drug therapy are increasingly important in decision
making. Clinical pharmacologists should welcome this as a means to promote efficiency and
effectiveness of prescribing, and aim to move the managers’ debate away from pure cost to the question of
value for money in prescribing.

VII. FURTHER READING

VII.1. Useful texts

   researchers. London: Dept of Health

VII.2. Useful introductory articles for nonspecialists

VIII. REFERENCES

Chapter 10. Drug Utilization

Folke Sjöqvist, M.D., Ph.D. ¹
Donald Birkett, Ph.D. ²

¹ Karolinska Institute
Division of Clinical Pharmacology
Huddinge University Hospital
S-141 86 Huddinge
SWEDEN

² Department of Clinical Pharmacology
Flinders Medical Centre
Flinders University of South Australia
Bedford Park, SA 5042
AUSTRALIA

Parts of this chapter are extracted from the WHO booklet “Introduction to Drug Utilization Research” 2003, ISBN 92 4 156234X, with the permission of the WHO Office of Publications (Mr. D.W. Bramley). The content in this chapter represents the views of the authors and not those of WHO. The authors were main contributors to the original WHO publication.
I. INTRODUCTORY REMARKS

WHAT IS DRUG UTILIZATION RESEARCH AND WHY IS IT NEEDED

I.1. Definitions

Drug utilization research was defined by WHO in 1977 as “the marketing, distribution, prescription, and use of drugs in a society, with special emphasis on the resulting medical, social and economic consequences”. Since then, a number of other terms have come into use and it is important to understand the interrelationships of the different domains. Epidemiology is defined as “the study of the distribution and determinants of health-related states and events in the population, and the application of this study to control of health problems”. Pharmacoepidemiology applies epidemiological methods to studies of the clinical use of drugs in populations. A suitable definition of pharmacoepidemiology is: The study of the use and effects/side effects of drugs in large numbers of people with the purpose of supporting the rational and cost-effective use of drugs in the population thereby improving health outcomes.

Pharmacosurveillance and pharmacovigilance are terms used to refer to the monitoring of drug safety such as spontaneous adverse effect reporting systems, case-control and cohort studies.

Pharmacoepidemiology may be drug-oriented, emphasizing the safety and effectiveness of individual drugs or groups of drugs, or utilization-oriented aiming to improve the quality of drug therapy through pedagogic intervention. Drug utilization research may also be divided into descriptive and analytical studies. The emphasis of the former has been to describe patterns of drug utilization and to identify problems deserving more detailed studies. Analytical studies try to link drug utilization data to figures on morbidity, outcome of treatment and quality of care with the ultimate goal being to assess whether drug therapy is rational or not. Sophisticated utilization-oriented pharmacoepidemiology may focus on the drug (e.g., dose-effect and concentration-effect relationships), the prescriber (e.g., quality indices of the prescription), or the patient (e.g., selection of drug and dose vs. kidney function, drug metabolic phenotype/genotype, age, etc).

Drug utilization research is thus an essential part of pharmacoepidemiology as it describes the extent, nature and determinants of drug exposure. In common use, the distinction between these two terms has become less sharp, and they are sometimes used interchangeably. However, while drug utilization studies often employ various sources of information focusing on drugs, e.g., aggregate data from wholesale and prescription registers, the term epidemiology implies defined populations and that drug use can be expressed in terms of incidence and prevalence.

Drug utilization research and pharmacoepidemiology may provide insights into the following aspects of drug use and drug prescribing:

Pattern of use: extent and profiles of drug use and trends in drug use and costs over time.

Quality of use: audits comparing actual use to national and regional prescription guidelines or local drug formularies. Quality indices of drug use may include the choice of drug (compliance to recommended assortment), drug cost (compliance to budgetary recommendations), drug dosage (awareness of inter-individual variations in dose requirements and age dependence), drug interaction

---

1 An audit in drug use was defined by Crooks (1979) as an examination of the way in which drugs are used in clinical practice carried out at intervals frequent enough to maintain a generally accepted standard of prescribing.
awareness, ADR awareness, proportion of patients being aware of/unaware of the cost/benefit of the treatment, etc.

**Determinants of use:** user characteristics (e.g. socio-demographic parameters, attitude towards drugs), prescriber characteristics (e.g. specialty, education and factors influencing therapeutic decisions), and drug characteristics (e.g. therapeutic properties, affordability)

**Outcomes of use:** health outcomes (benefits and adverse effects) and economic consequences.

Pharmacoepidemiology initially focused on the safety of individual drug products (pharmacosurveillance), but now also includes studies of their beneficial effects. The driving force behind this development was a growing awareness that health outcomes of drug use in the rigorous setting of randomized clinical trials is not necessarily the same as health outcome of drug use in everyday practice. The clinical trials that are needed to obtain marketing authorization for new drugs involve limited samples of carefully selected patients, who are treated and followed-up for a relatively short period of time in strictly controlled conditions. As a result, such trials do not provide an accurate reflection of how drug use will impact health outcomes in everyday practice under everyday circumstances. Pharmacoepidemiological studies often make useful contributions to our knowledge about effectiveness and safety, because they assess drug effects in large, heterogeneous patient populations over longer periods.

Drug utilization research also provides insight into the efficiency of drug use, i.e. whether a certain drug therapy provides value for money. Drug utilization research can thus help to set priorities for the rational allocation of health care budgets.

I.2. Why drug utilization research?

The principal aim of drug utilization research is to facilitate rational use of drugs in populations. For the individual patient rational use of a drug implies the prescription of a well-documented drug in an optimal dose on the right indication, with the correct information and at an affordable price. Without knowledge on how drugs are being prescribed and used, it is difficult to initiate a discussion on rational drug use and to suggest measures to change prescribing habits for the better. Information on the past performance of prescribers is the linchpin of any auditing system.

Drug utilization research in itself does not necessarily provide answers, but it contributes to rational drug use in three important ways:

I.2.A. Description of drug use patterns

Drug utilization research will increase our understanding of how drugs are being used by:

a. Making estimates of the numbers of patients exposed to drugs within a given time period. Such estimates may either refer to all drug users, regardless of when they started to use the drug (prevalence), or focus on patients who started to use the drug within the selected period (incidence).

b. Describing the extent of use at a certain moment and/or in a certain area (e.g. country, region, community, hospital). Such descriptions are most meaningful when they are part of a continuous evaluation system, i.e. when the patterns are followed over time and trends in drug use can be described.

c. Estimating (e.g. on the basis of epidemiological data on a disease) to what extent drugs are properly used, overused, or underused.

d. Describing the pattern or profile of drug use - assessing which alternative drugs are being used for particular conditions and to what extent.
Comparing observed patterns of drug use with current recommendations or guidelines for the treatment of a certain disease.

Applying quality indicators to drug utilization patterns. An example is the so-called DU90% (drug utilization 90%), a further development of the “Top-10” list. The DU90% segment reflects the number of drugs that account for 90% of drug prescriptions and adherence to local or national prescription guidelines in this segment. This general indicator can be applied at different levels (individual prescriber, group of prescribers, hospitals, region, county, etc.) to get a rough estimate of the quality of prescribing.

Feeding back drug utilization data to prescribers. This is particularly useful when the individual’s drug prescribing can be compared with some form of "gold standard" or best practice, and with the average prescriptions in the country, the region, or the area.

Relating the number of case reports about a drug problem or adverse effects to the number of patients exposed in order to assess the potential magnitude of the problem. If it is possible to detect that the reaction is more common in a certain age group, in certain conditions or at a special dose level, improving the information on proper use such as indications, contraindications and appropriate dosages may be sufficient to assure a safer use. Thereby withdrawal of the drug from the market may be avoided.

I.2.B. Early signals of irrational use of drugs

Drug utilization research may generate hypotheses that set the agenda for further investigations by:

a. Comparing drug utilization patterns and costs between different regions or time periods. Hypotheses can be generated to form the basis for investigations of the reasons for, and health implications of, the differences found. Geographical differences and changes over time in drug use may have medical, social and economic implications both for the individual patient and for society, and are thus important to identify, explain and sometimes correct.

b. Comparing observed patterns of drug use with current recommendations/guidelines for the treatment of a certain disease. Hypotheses can then be generated about whether discrepancies represent less than optimal practice, whether pedagogic interventions (education) are required, or whether the guidelines need to be reviewed in the light of actual practice. These considerations should include both underuse and overuse of drugs.

I.2.C. Interventions to improve drug use – follow-up

Drug utilization research may enable us to assess whether interventions undertaken to improve drug use have had the desired impact by:

a. Monitoring and evaluating the effects of measures taken to improve undesirable patterns of drug use (regional or local formularies, information campaigns, regulatory policies, etc.)

b. Following the impact of regulatory changes or changes in insurance or reimbursement systems. This also requires a broad survey, because the total cost to society may remain the same or may even increase, if other more expensive drugs are used as an alternative.

c. Assessing to which extent promotional activities of the pharmaceutical industry and educational activities of the society impact on the patterns of drug use.

II. TYPES OF DRUG USE INFORMATION

Different types of drug use information are required depending on the problem being evaluated. These include information about the overall drug use, or use of drug groups, individual generic compounds or specific products. Often, information about the condition being treated, about the patient and about the
prescriber will be required. In addition, data on drug costs will be important in ensuring that drugs are used efficiently and economically. These types of drug information are described below.

**II.1 Drug based information**

The trends in total drug use may sometimes be useful to know, but more detailed information is usually required to answer clinically important questions. This may involve aggregation of drug use at various levels, and information on indications, doses and dosage regimens.

**II.2. Problem or encounter-based information**

Instead of asking how a particular group of drugs is used, one may well address the question how a particular problem (e.g. sore throat, hypertension, gastric ulcer, depression) is managed.

**II.3. Patient information**

Demographic and other information about the patient will often be useful. The age distribution of patients will sometimes be of critical importance, for example to assess the likelihood of severe adverse effects with NSAIDs, or whether the drug is being used in an age group different to that in which the clinical trials were performed. The co-morbidities of the patient group may be important in determining treatment choice and adverse effects. As an example in the management of hypertension, beta-blockers should be avoided in patients with asthma, and ACE inhibitors preferred in patients with heart failure.

Qualitative information such as knowledge, beliefs, and perceptions among patients and their attitudes to drugs will be important in some cases, for example in assessing patient pressures on doctors to prescribe antibiotics, or in designing consumer information/education programs.

**II.4. Prescriber information**

The prescriber is a critical point in determining drug use. Some sceptics even claim that doctors differ more than patients and that differences in drug prescribing often lack rational explanations. Dissecting the factors that determine prescribing behaviour is therefore often central to understanding how and why drugs are prescribed.

**III. SOURCES OF DRUG UTILIZATION DATA**

The drug use chain includes the processes of drug acquisition, storage, distribution, prescribing, patient compliance and review of outcome of treatment. Each of these events is an important aspect of drug utilization. Drug utilization data may be derived from quantitative or qualitative studies. Quantitative data may be used to describe the present state, and trends in drug prescribing and drug use at various levels of the health care system. Quantitative data are usually obtained from routinely collected data or from surveys. Qualitative studies assess the appropriateness of drug utilization and generally link prescribing data to reasons (indications) for prescribing. Such studies have been referred to as Drug Utilization Review or Drug Utilization Evaluation. The process is one of a “therapeutic audit” based on defined criteria and has the purpose of improving the quality of therapeutic care. There is an increasing interest in the evaluation of the economic impact of clinical care and medical technology. This has evolved into a discipline dedicated to the study of how pharmacotherapeutic methods influence resource utilization in health – pharmacoeconomics.

The increasing interest in efficient use of health care resources has resulted in the establishment of computer databases for studies on drug utilization. Some of the databases can generate statistics for
patterns of drug utilization and adverse drug reactions. Data may be in the form of drug sales, drug movement at various levels of the drug distribution chain, pharmaceutical and medical billing data or samples of prescriptions. Data may also be obtained from drug importers, wholesalers or local manufacturers.

Data from medical practices and health facilities may be used to measure specific aspects of health provision and drug use. Such data may be used to generate indicators that provide information on prescribing habits and aspects of patient care. These indicators can be used to determine where drug use problems exist, provide a mechanism for monitoring and supervision and motivate health care providers to follow established health care standards.

Prescription and dispensing data are useful for determining some of the quality indicators of drug use recommended by WHO. These include:

- Average number of drugs per prescription (encounter)
- Percentage of drugs prescribed by generic name
- Percentage of encounters with an antibiotic prescribed
- Percentage of encounters with an injection prescribed
- Percentage of drugs prescribed from essential drugs list or formulary
- Average drug cost per encounter

III.1 Drug use evaluation

Drug use evaluation is a system of ongoing, systematic, criteria-based drug evaluation that ensures the appropriate use of drugs. Drug use evaluation is sometimes referred to as drug utilization review. It is a method of obtaining information to identify problems of drug use. Properly developed, it not only provides a means of identifying drug use problems but also provides a means to correct the problem and thereby contributes to rational drug therapy.

Drug use evaluation can assess the actual process of medication administration or dispensing (appropriate indications, drug selection, dose, route of administration, duration of treatment, drug interactions) and also assess outcomes of treatment (cured disease conditions, decreased levels of a clinical parameter). The objectives of drug use evaluation include:

- Ensuring that drug therapy meets current standards of care
- Controlling drug cost
- Preventing medication related problems
- Evaluating the effectiveness of drug therapy
- Identification of areas of practice that require further education of practitioners

Identification of problems to be subjected to drug use evaluation may be obtained from any of the data from the practice setting section (prescription indicators, dispensing data, aggregate data). The main source of data for drug use evaluation is the patient records. An identifiable authoritative group, like the Drugs and Therapeutic Committee, usually carries out drug use reviews in the hospital or health facility. This group has the responsibility of drawing up the guidelines, criteria, indicators and thresholds for the evaluation. Drug use evaluation may be based on data collected prospectively (as drug is being dispensed or administered) or retrospectively (based on chart reviews or other data sources).

IV. DRUG CLASSIFICATION SYSTEMS

A drug classification system represents a common language for describing the drug assortment in a country or region and is a prerequisite for national and international comparisons of drug utilization
data, which have to be collected and aggregated in a uniform way. Access to standardised and validated information on drug use is essential to allow audits of patterns of drug utilization, to identify problems in drug use, to initiate educational or other interventions and to monitor the outcomes of the interventions. The main purpose of having an international standard is to be able to compare data between countries. A recent example is the international focus on creating comparable monitoring systems for cross-national antibacterial utilization patterns in the work against bacterial resistance.

IV.1 Different classification systems

Drugs can be classified in different ways: according to their mode of action, according to indications, or according to chemical structure. Each classification system will have its advantages and limitations and the usefulness will depend on the purpose, the setting used, and the user’s knowledge of the methodology. Comparisons between countries may require a different classification system than a local comparison (e.g. between different wards in a hospital). Of the various systems proposed over the years, only two have survived to attain a dominant position in drug utilization research worldwide. These are the Anatomical Therapeutic Classification (AT) developed by the European Pharmaceutical Market Research Association (EPhMRA) and the Anatomical Therapeutic Chemical (ATC) classification developed by Norwegian researchers. These systems were originally based on the same main principles. In the EPhMRA system, drugs are classified in groups at three or four different levels. The ATC classification system is modified and extended from the EPhMRA system by the addition of a therapeutic/pharmacological/chemical subgroup as the fourth level and the chemical substance as the fifth level.

ATC is also the basis for the classification of adverse drug reactions used by the WHO Collaborating Centre for International Drug Monitoring in Uppsala (www.who-umc.org).

The main purpose of the ATC classification is as a tool for presenting drug utilization statistics and it is recommended by the WHO to be used in international comparisons. The EPhMRA classification system is used worldwide by IMS (Intercontinental Medical Statistics) for providing marketing research statistics to the pharmaceutical industry. It should be emphasised that there are many technical differences between the EPhMRA classification and the ATC classification. Therefore, data prepared using the ATC classification cannot be directly compared with those obtained with the EPhMRA system. In 1996, WHO established the ATC/DDD system as an international standard in drug utilization studies.

IV.2. The ATC classification system

In the Anatomical Therapeutic Chemical classification system the drugs are divided into different groups according to the organ or system on which they act and their chemical, pharmacological and therapeutic properties.

Drugs are classified in groups at five different levels. The drugs are divided into fourteen main groups (1st level), with two therapeutic/pharmacological subgroups (2nd and 3rd levels). The 4th level is a therapeutic/pharmacological/chemical subgroup and the 5th level is the chemical substance. The 2nd, 3rd and 4th levels are often used to identify pharmacological subgroups when this is considered to be more appropriate than therapeutic or chemical subgroups.

The complete classification of glibenclamide illustrates the structure of the code:
Thus, in the ATC system all plain glibenclamide preparations are given the code A10B B01.

Alterations in the ATC classification are made when the main use of a drug has clearly changed, and when new groups are required to accommodate new substances or to achieve better specificity in the groupings.

In the ATC system drugs are separated into groups at five different levels (described above). By use of this classification system, statistics of drug utilization grouped at five different levels can be provided; from figures showing total drug use of all products classified e.g. in a main group (1st level), to figures for the different subgroups (2nd, 3rd and 4th level) and down to figures showing use of the separate substances.

The publication *Guidelines for ATC Classification and DDD Assignment* gives further and detailed information about the ATC classification. (WHO Collaborating Centre for Drug Statistics Methodology, 2003; www.whocc.no)

V. DRUG UTILIZATION METRICS AND THEIR APPLICATIONS

V.1. The concept of the defined daily dose (DDD)

The DDD is the assumed average maintenance dose per day for a drug used for its main indication in adults.

The defined daily dose is a unit of measurement and does not necessarily agree with the recommended or prescribed daily dose (PDD). Doses for individual patients and patient groups will often differ from the DDD and have to be based on individual characteristics (e.g. age and weight) as well as pharmacokinetic and pharmacogenetic considerations.

The DDD is often a compromise based on a review of the available information about doses used in various countries. The DDD may even be a dose that is rarely prescribed, because it is an average of two or more commonly used dose sizes.

Drug utilization figures should preferably be presented as numbers of DDDs/1000 inhabitants/day or, when in-hospital drug use is considered, as DDDs per 100 bed days. For antiinfectives (or other drugs normally used in short periods), it is often considered most appropriate to present the figures as numbers of DDDs per inhabitant per year.
These terms are explained in the following:

V.1.A. DDDs/1000 inhabitants/day
Sales or prescription data presented in DDD/1000 inhabitants/day may provide a rough estimate of the proportion of the study population that may be treated daily with certain drugs. As an example, the figure 10 DDDs/1000 inhabitants/day indicates that 1% of the population on average might get a certain drug or group of drugs every day. This estimate is most useful for chronically used drugs when there is good agreement between the average prescribed daily dose (see below) and the DDD. It may also be important to consider the size of the population used as a denominator. Usually the general utilization is calculated for the total population including all age groups. Some drug groups have very limited use among young people, with most users above the age of 45. To correct for utilization differences due to differing age structures between countries, simple age adjustments can be made by using the number of inhabitants in the relevant age group as a denominator.

V.1.B. DDDs per 100 bed days
This unit may be applied when in-hospital drug use is considered. As an example, 70 DDD/100 bed days of hypnotics provide an estimate of the therapeutic intensity and suggests that 70% of the in-patients might receive a DDD of a hypnotic every day. This unit is quite useful for benchmarking in hospitals.

V.1.C. DDDs per inhabitant per year
This term may give an estimate of the number of days for which each inhabitant is, on average, treated annually. For example, 5 DDDs/inhabitant/year indicates that the utilization is equivalent to the treatment of every inhabitant with a 5 days course during a certain year. Alternatively, if the standard treatment period is known, the total number of DDDs can be calculated as the number of treatment courses, and the number of treatment courses can then be related to the total population.

V.2. Prescribed daily dose/Consumed daily dose
The prescribed daily dose (PDD) is defined as the average dose prescribed according to a representative sample of prescriptions. The PDD can be determined from prescription studies and medical- or pharmacy records. It is important to relate the PDD to the diagnosis for which the dosage is based. The PDD will give the average daily amount of a drug that is actually prescribed. When there is a substantial discrepancy between the PDD and the defined daily dose (DDD), it is important to take this into consideration when evaluating and interpreting drug utilization figures, particularly in terms of morbidity.

The PDD can vary according to both the illness treated and national therapeutic traditions. There are also substantial differences between PDDs in various countries. PDDs in Asian populations are often lower than in Caucasians. The fact that PDDs may differ from one country to another should always be considered in international comparisons.

It should be noted that the prescribed daily dose does not necessarily reflect actual drug utilization. Some prescribed medications are not dispensed, and the patient does not always take all the medications that are dispensed. Specially designed studies including patient interviews are required to measure actual drug intake at the patient level (i.e. consumed daily dose).

V.3. Volume
Common physical units (e.g. grams, kilos, litres), numbers of packages or tablets and numbers of prescriptions are also used for quantifying drug utilization. These units can be applied only when the use of one drug or of well-defined products is evaluated.
Pharmacological Research in Cardiovascular Disorders

Chapter 11. Hypertensive Vascular Disease

Chapter 12. Lipid Lowering Agents

Chapter 13. Anti-atherosclerotic Drugs

Chapter 14. Heart Failure

Chapter 15. Arrhythmias
Chapter 11. Hypertensive Vascular Disease

Pierre Larochelle, M.D., Ph.D. 1

Maxime Lamarre-Cliche, M.D. 2

1 Directeur de la recherche clinique
Unité de recherche en pharmacologie cardiovasculaire
Institut de recherche clinique de Montréal
110, avenue des Pins Ouest
Montréal, Québec H2W 1R7
CANADA

2 Chargé de recherche
Unité de recherche en pharmacologie cardiovasculaire
Institut de recherche clinique de Montréal
110, avenue des Pins Ouest
Montréal, Québec H2W 1R7
CANADA
I. INTRODUCTORY REMARKS

Over the last 3 decades, the treatment of essential hypertension has evolved with new objectives and new classes of drugs. The objectives of the treatment have changed from the reduction in blood pressure to the prevention of cardiovascular disease by the reduction and normalization of blood pressure. Treatment is also evolving towards a more individual approach of the treatment taking into considerations such variables as age, gender, race and the associated diseases such as among others diabetes, renal function, and cardiac function. There have been new classes of drugs which have been evaluated such as the calcium channel blockers, the ACE inhibitors, the renin-angiotensin blockers and others which are currently being tested such as the aldosterone receptor blockers.

The inhibition of the renin-angiotensin system was successfully used in the treatment of hypertension and heart failure. The first approach used was the inhibition of the angiotensin-converting enzyme (ACE). However, ACE-inhibitors induce the accumulation of other peptides such as substance P or bradykinin, and consequently untoward drug effects like cough and angioedema can become apparent. A more recent approach to counteract increased blood pressure and sympathetic tone, as well as harmful cardiovascular hypertrophy and renal lesions, was to decrease the activity of angiotensin II receptors. Pharmacological blockade of AT₁-subtype angiotensin II receptors appears to be clinically equally effective but the generation of angiotensin II remains unopposed during AT₁-blockade and leaves the potential for stimulation of other angiotensin II receptor subtypes. These two classes of drugs have been proven to be effective and very well tolerated. These drugs have not been shown to be superior to conventional treatment such as diuretics in a general population but have been shown to be superior to beta blockers in some groups of patients with hypertension such as the patients with LVH (left ventricular hypertrophy). Despite the addition of these two classes of drugs, there is still definitely a need for more effective, as safe and possibly more specific drugs to treat hypertension. Moreover, the effect of antihypertensive therapy on cardiovascular events, on cerebrovascular events, on diabetic complications progression needs to be part of the evaluation process. Further studies are needed to document the long-term benefit of antihypertensive therapy alone or in association with other cardiovascular medications.

II. PHASE II STUDIES TO EVALUATE THE EFFECTIVENESS AND SAFETY OF NEW ANTIHYPERTENSIVE DRUGS

Studies will have to assess the effect of a drug directly on blood pressure, e.g. looking primarily for an antihypertensive action and they will have also to assess the effect of the drug in special situations such as renal dysfunction or others and in special populations.

Long term studies may have eventually as an objective to assess the effect of an antihypertensive agent on cardiovascular events such as coronary events, including myocardial infarction, sudden or rapid death from cardiac causes, heart failure, acute occlusion of a major feeding artery in any vascular bed other than cerebral or coronary, death from non-coronary cardiac causes, dissecting or ruptured aortic aneurysm, or death from vascular causes, cerebrovascular events, including stroke and transient ischemic attacks, and on progression of hypertensive kidney disease, diabetic nephropathy, diabetic retinopathy, and arterial wall thickness. These studies usually are conducted over the period of 5 years and are considered long-term studies.

II.1. Outline of a potential development plan

Multicenter, randomized, double-blind, placebo-controlled, parallel group study in patients with mild-to-moderate essential hypertension (WHO classification grades 1 and 2) who have been completely withdrawn from their previous antihypertensive medication or in patients who have been newly diagnosed with mild-to-moderate essential hypertension and who are not currently taking any antihypertensive
medication(s). The objective of these phase II studies is to find the effective dose of the compound compared to a placebo. These are normally dose finding placebo controlled studies. A reference drug can be included in some studies. These studies are of primary importance to determine the dose which will be used in the phase III studies to demonstrate better efficacy and tolerability than the marketed compounds.

In these phase II protocols or in phase III protocols, the evaluation of the antihypertensive drugs in special situations can be assessed. These studies need to give the information of the effect of the drug in the elderly whether defined as over 65 or 75 years. Information needs to be obtained in both groups. More and more information will be asked in the individuals over the age of 85 years. There is limited data on the benefit of treatment in this age category but we need also at least to obtain information on the effectiveness of the drug in terms of BP lowering and dose efficacy.

Protocols need to be done in patients with isolated systolic hypertension or primarily systolic hypertension. These patients are generally elderly individuals with a specific form of hypertension which is becoming more prevalent as our population ages.

Information is required on the use of the drug according to gender and race. Black population may have a different response than other races as has been shown now in various studies.

Specific information needs also to be obtained in subgroups of hypertensive patients. Those subgroups are those with renal dysfunction that is, with a creatinine clearance of <60 ml/min, <30 ml/min and lower, as well as the patients with hepatic dysfunction

Special populations which need to be studied are also those with associated disease such as diabetes and left ventricular hypertrophy. Although not a requirement, this information will become necessary to the proper utilization of the drug

II.2. Short term studies

II.2.A. Study Objectives

Primary objectives
a. To determine the efficacy of the investigational drug X at given doses compared to placebo or the reference drug Y in patients with WHO classification grades 1 and 2 uncomplicated diastolic essential hypertension (mean sitting diastolic blood pressure [MSDBP] ≥ 95 mmHg and < 110 mmHg).
b. To determine the safety of the investigational drug X compared to the placebo in patients with WHO classification grades 1 and 2 uncomplicated diastolic essential hypertension (MSDBP ≥ 95 mmHg and < 110 mmHg).

Secondary objectives
a. To determine the efficacy and safety of different doses of the investigational drug X compared to placebo in the treatment of patients with WHO classification grades 1 and 2 uncomplicated systolic essential hypertension (systolic blood pressure [MSSBP] ≥ 145 mmHg and < 180 mmHg).
b. To assess the effects of the investigational drug X and the reference drug on standing blood pressure, sitting pulse and standing pulse.

II.2.B. Primary endpoints
a. Change from baseline in MSDBP at trough.
Response rate: patients were defined as "normalized" responders if their blood pressure values were < 140/90 mm Hg, or as "non-normalized" responders if the reduction in blood pressure was 10 mm Hg or more, compared with baseline.

b. Changes in sitting blood pressure at the end of the study.

II.2.C. Secondary endpoints
a. Change from baseline in MSSBP at trough.

b. Other variables to be analyzed include the change in standing diastolic and systolic blood pressures, sitting and standing pulse.

A variable such as the ambulatory blood pressure measurement (ABPM) is a requirement now to more precisely measure the course of action of the medication over the period of 24 hours including the peak effect and the trough effect. It will help determine the duration of action, the peak to trough ratio. Specific protocols will be planned to evaluate the effects of the drug on ABPM and some of its variables (day pressure, night pressure, smoothness of curve etc)

II.2.D. Study Design
Eligible patients enter a washout period during which antihypertensive medication is withdrawn and no other is allowed. The washout period is followed by a 2 to 4 -week single-blind placebo run-in period. Patients who meet the study inclusion/exclusion criteria at the end of the single-blind placebo run-in period are then randomized in double-blind fashion to either the investigational drug X or placebo once daily for an 8-week treatment period in a parallel designed trial. The effect of the investigated medication can also be evaluated at the end of the trial by a period of drug withdrawal by recording blood pressure and adverse events from one to seven days or more after the last dose of study medication.

Blood pressure measurement
Blood pressure will be measured using a calibrated standard mercury sphygmomanometer or a calibrated electronic automatic sphygmomanometer with digital reading. Personnel recording the blood pressure should receive training on the measurement of blood pressure and follow guidelines such as the one reported in the annex 1. The standard measurement is done in the sitting position, but standing pressure is also recorded in most protocols, as well as pulse rate. Sitting and standing blood pressure will be measured and recorded at each visit. Every effort will be made to have the same staff member obtain blood pressure measurements for the same patient, at the same time of day, using the same equipment, at each visit.

Pulse rate
At each visit, the pulse rate will be measured for 30 seconds just prior to the blood pressure measurements; once in the sitting position and once in the standing position.

Concomitant therapy
Use of the following medications may interfere with the evaluation of efficacy, safety and/or tolerability. Therefore, these medications are excluded throughout the trial, from the beginning of the washout period until the end of the double-blind treatment period. Patients who are receiving such medications should be excluded, or if ethically justified, the medications may be withdrawn according to the manufacturer's instructions.

a. Drugs approved for the treatment of hypertension even if prescribed for another indication.

b. Any antidepressant drugs in the class of MAO inhibitors and tricyclics. Other psychotrophic drugs such as benzodiazepines and selective serotonin reuptake inhibitors (SSRIs) will be allowed if well tolerated when previously taken, and if the dosage is expected to remain stable throughout the study.
c. Any systemic use of corticosteroids. Topical and nasal steroid preparations may be used as needed, but not on a daily basis.
d. Use of hormonal contraceptives, including subdermal contraceptive implants, within one month prior to Visit 1 (Week -4).
e. Thyroid medication and/or oestrogen replacement therapy, unless these have been stable maintenance replacement doses for 6 months preceding Visit 1 (Week -4).
f. Insulin.
g. Chronic administration of sympathomimetic drugs such as those found in nasal decongestants (pseudoephedrine and phenylpropanolamine) and bronchodilators (e.g., metaproterenol).
h. Antacids in amounts greater than package labelling.
i. Ergot preparations.
j. Antiarrhythmic drugs; digoxin will be permitted provided serum levels have been stable and no dose adjustments have been made during the 6 months preceding Visit 1 (Week -4).
k. Diuretics of any kind.
l. Psychotropic drugs, except for hypnotics and mild anxiolytic agents such as benzodiazepines, if these were used occasionally (pm) before the start of the study.
m. Aspirin above 325 mg daily. Aspirin will be allowed at a maximum daily dose of 325 mg for cardiac protection. Patients must have been on a stable dose prior to entry and maintain the same dose throughout the study.
n. The use of drug(s) in the 6 months prior to Visit 1 (Week -4) which are potentially hepatotoxic (e.g., methotrexate) or nephrotoxic (e.g., gentamicin).
o. Antianginal medication of any kind including calcium channel blockers or beta blockers (including beta blocker eye drops).
p. Drugs which interfere with the metabolism of the other compound such as through the inhibition or stimulation of isozymes of the cytochrome P450 known to influence the compound under study should be excluded.

II.2.E. Planned sample
The sample size is calculated on the basis of the objective to be obtained. Sample size to obtain a clinically significant reduction or difference in blood pressure can be calculated with sufficient power for example. Statistical help is required for the best accurate determination of these calculations and different tools are available on the internet to calculate sample size:
   a. To conclude non-inferiority of the investigational drug within a margin of 2 mm Hg when there is no true treatment difference between the investigational drug X mg and the reference drug Y mg OD.
   b. To detect a true 3 mm Hg difference between the reference drug Y mg (or the investigational drug X mg) OD and placebo under the null hypothesis that the mean difference is 0.
   c. To have sufficient power to examine dose-response relationship of the investigational drug at different dose levels.

II.2.F. Study population
Patients with mild-to-moderate essential hypertension (WHO classification grades 1 and 2) who have been completely withdrawn from their previous antihypertensive medication or in patients who have been newly diagnosed with mild-to-moderate essential hypertension and who are not currently taking any antihypertensive medication(s).

II.2.G. Inclusion criteria
   a. Outpatients 21 and older. The age limit will be determined by the protocol or the planned development program which should include at one point studies in elderly and very elderly patients
   b. Male or female patients are eligible. Female patients must be either post-menopausal for one year or surgically sterile, or using effective contraceptive methods such as barrier method with
spermicide or an intra-uterine device. Oral or implant contraceptives are not be allowed because of their interactions and effect on blood pressure.

c. Patients with mild to moderate essential diastolic hypertension (WHO classification grades 1 or 2) measured by calibrated standard sphygmomanometer or calibrated standardized automatic sphygmomanometer. Patients must have a MSDBP ≥ 95 mmHg and < 110 mm Hg. Blood pressure criteria will be different in studies such as those in diabetic patients.

d. Patients must have a variability of ≤ 10 mmHg in their MSDBP between pre-randomization.

e. Patients who are eligible and able to participate in the study, and who consent to do so after the purpose and nature of the investigation have been clearly explained to them (written informed consent).

II.2.H. Exclusion criteria

Patients with any of the following physiological states or concomitant medical conditions will be excluded from further participation in the study.

a. Severe hypertension (defined as MSDBP ≥ 110 mmHg and/or MSSBP ≥ 180 mmHg or 200 mm Hg).

b. Inability to discontinue all prior anti-hypertensive medications safely for a period of 12 weeks.

c. Known Keith-Wagener grade III or IV hypertensive retinopathy.

d. History of hypertensive encephalopathy or cerebrovascular accident within the preceding 6 months.

e. Transient ischemic cerebral attack during the 12 months prior to Visit 1 (Week -4).

f. Evidence of a secondary form of hypertension, such as coarctation of the aorta, hyperaldosteronism, unilateral renal disease, or pheochromocytoma, etc.

g. Type 1 diabetes mellitus.

h. Type 2 diabetes mellitus with poor glucose control as defined by fasting glycosylated hemoglobin (HbA1c) >8% or requiring insulin treatment.

i. Known or suspected contraindications, including history of hypersensitivity to the class of antihypertensive agent.

j. Any surgical or medical condition which might significantly alter the absorption, distribution, metabolism, or excretion of any drug including but not limited to any of the following:

   • History of major gastrointestinal tract surgery such as gastrectomy, gastroenterostomy, or bowel resection.

   • Currently active or previously active inflammatory bowel syndrome during the 12 months prior to Visit 1 (Week -4).

   • Currently active gastritis, duodenal or gastric ulcers, or gastrointestinal/rectal bleeding during the 3 months prior to Visit 1 (Week -4).

   • Any history of pancreatic injury, pancreatitis or evidence of impaired pancreatic function/injury as indicated by abnormal lipase or amylase.

   • Evidence of hepatic disease as determined by any one of the following: SGOT or SGPT values exceeding 2 x ULN at Visit 1 (Week -4), a history of hepatic encephalopathy, a history of oesophageal varices, or a history of portocaval shunt.

   • Evidence of renal impairment as determined by any one of the following: serum creatinine > 1.5 ULN or a level higher than normal depending on the population under study, a history of dialysis, or a history of nephrotic syndrome.

   • Current obstruction of the urinary tract or difficulty in voiding due to mechanical or inflammatory conditions which is likely to require intervention during the course of the study or is regarded as clinically meaningful by the investigator.

k. History or diagnosis of heart failure during the 6 months prior to Visit 1 (Week -4).

l. History of myocardial infarction during the 6 months prior to Visit 1 (Week -4).

m. Second or third degree heart block without a pacemaker.

n. Unstable angina pectoris.

o. Concurrent potentially life threatening arrhythmia or symptomatic arrhythmia.
p. Clinically significant valvular heart disease.
q. Volume depletion or dehydration.
r. History of malignancy including leukemia and lymphoma (but not basal cell skin cancer) within the past five years.
s. History of any severe or life-threatening disease(s).
t. Any surgical or medical condition, which in the opinion of the investigator, may place the patient at higher risk from his/her participation in the study, or is likely to prevent the patient from complying with the requirements of the study or completing the trial period.
u. History of drug or alcohol abuse within the last 12 months.
v. History of non-compliance to medical regimens or unwillingness to comply with a study protocol.
w. Participation in any investigational drug trial within one month of Visit 1 (Week -4).
x. Unwillingness or inability to give informed consent.
y. Persons directly involved in the execution of this protocol.
z. History of any severe or life-threatening disease(s).

II.2.I. Tools for assessing endpoints

Efficacy assessment
Using a calibrated standard sphygmomanometer and appropriate size cuff, arterial blood pressure determinations will be made in accordance with the Seventh report of the Joint National Committee on Prevention, Detection, Evaluation, and Treatment of High Blood Pressure or that of other societies (1). With the arm supported at the level of the heart, systolic pressure will be recorded when the initial sound is heard (Phase I of the Korotkoff sound); diastolic pressure will be recorded at the disappearance of the sound (Phase V of the Korotkoff sound). The cuff should be deflated at a rate not greater than 2 mm Hg/sec.

The ambulatory blood pressure measurement (ABPM) will be measured using standard procedures (Appendix 2).

Safety assessments
Safety assessments will consist of monitoring and recording all adverse events (AEs) and serious adverse events (SAEs), the regular monitoring of haematology, blood chemistry and urine values, regular measurement of vital signs and the performance of physical examinations. An ECG evaluation will be obtained at Visit 2 (Week -2).

II.2.J. Specific criteria for early withdrawal and discontinuation

a. Reasons why a patient may discontinue participation in a study include the following:
b. Adverse event(s)
c. Abnormal laboratory value(s)
d. Abnormal test procedure result(s)
e. Unsatisfactory therapeutic effect
f. Patient's condition no longer requires study treatment
g. Protocol violation
h. Subject withdrew consent
i. Lost to follow-up
j. Administrative reasons
k. Death
Patients with MSDBP $\geq 110$ mmHg or MSSBP $\geq 180$ mmHg at any time during the single-blind or double-blind treatment phases must be permanently discontinued from the trial.

II.2.K. Data analysis method
The proportion of patients in each treatment achieving a successful reduction in MSDBP during the double-blind period will be compared at the endpoint using a one-way logistic model with treatment as the factor at Visit 7 (Week 8) for all randomized patients. Success is defined as a mean sitting diastolic blood pressure < 90 mmHg or a decrease $\geq 10$ mmHg sitting diastolic blood pressure from the randomization visit.

The changes from baseline will be analyzed using a two-way analysis of covariance model with treatment and center as factors, and the baseline as a covariate as well as treatment-by-baseline interaction. All pairwise treatment comparisons will be made based on this analysis model. Due to a large number of study centers and treatment groups planned in this study, treatment-by-center interaction effect may be difficult to interpret in a statistical model. However, a summary of means by treatment and center will be provided for the primary analysis.

III. PHASE III STUDIES TO EVALUATE THE EFFECTIVENESS AND SAFETY OF NEW ANTIHYPERTENSIVE DRUGS

III.1. Outline of a typical development plan
The phase III studies have the objective to demonstrate and compare the efficacy of the antihypertensive agent against the usual antihypertensive compounds. The new compound will be tested for efficacy and compared to conventional treatments such as diuretics (hydrochlorothiazide), beta blockers (metoprolol or atenolol), long acting dihydropyridine calcium channel blockers (amlodipine or long acting nifedipine), ACE inhibitors (enalapril or lisinopril) and AT1 receptor blockers such as losartan.

Since the treatment of hypertension requires more than one antihypertensive agent, it will also be a requirement to test the new agent in combination to the most commonly used other antihypertensive agent either as a specific trial such as a combination with a thiazide diuretic or as an add-on protocol where non responsive patients will receive add on treatments to normalize blood pressure. Combination will also be required where the new medication will be added as a second line agent to conventional treatment to evaluate potential synergistic effect on BP reduction but also potential adverse effects.

Phase III long-term studies are done to follow up on the long term effectiveness on blood pressure and adverse effects of the medication over a period which could be up to two years of follow up in small number of patients who generally have participated in phase II or III studies and are asked to continue on the medication in an open study of efficacy and safety.

III.2. Short-term studies

III.2.A. Objectives
The objective is to demonstrate the relative efficacy and safety of the new compound.

III.2.B. Primary endpoints
a. The decrease in blood pressure whether systolic or diastolic blood pressure measured in standard conditions at through compared to other antihypertensive treatments.

b. Responders rate are also measured in terms of % of patients obtaining a predetermined endpoint of BP such as diastolic BP less than 90 mm Hg and systolic pressure less than 160 mm Hg or 140 mm Hg.
III.2.C. Secondary end-points
As described for phase II studies in II.2.C.

III.2.D. Study design
Eligible patients will enter a 2 to 4 weeks washout period followed by a 2 to 4 weeks single blind placebo period to determine if patients can meet the inclusion criteria for blood pressure or if they have to be excluded for other exclusion criteria. They will then be randomized to either treatments conventional or experimental to evaluate efficacy and safety. The duration of the short term active treatment phase is of the order of 8 weeks.

III.2.E. Planned sample
As described for phase II studies in II.2.E.

III.2.F. Study population
As described for phase II studies in II.2.F.

III.2.G. Specific inclusion criteria
As described for phase II studies in II.2.G.

III.2.H. Specific exclusion criteria
As described for phase II studies in II.2.H.

III.2.I. Tools for assessing endpoints
As described for phase II studies in II.2.I.

III.2.J. Specific criteria for early withdrawal and discontinuation
As described for phase II studies in II.2.J.

III.2.K. Data analysis method
The proportion of patients in each treatment achieving a successful reduction in MSDBP during the double-blind period will be compared at the endpoint using a one-way logistic model with treatment as the factor at Visit 7 (Week 8) for all randomized patients. Success is defined as a mean sitting diastolic blood pressure < 90 mmHg or a decrease ≥10 mmHg sitting diastolic blood pressure from the randomization visit.

The changes from baseline will be analyzed using a two-way analysis of covariance model with treatment and center as factors, and the baseline as a covariate as well as treatment-by-baseline interaction. All pairwise treatment comparisons will be made based on this analysis model. Due to a large number of study centers and treatment groups planned in this study, treatment-by-center interaction effect may be difficult to interpret in a statistical model. However, a summary of means by treatment and center will be provided for the primary analysis.

The following pairwise comparisons shall be performed:
   a. Investigational vs. reference drug
      * Primary comparison: non-inferiority and/or superiority of the investigational drug X mg OD vs. reference drug Y mg OD.
         i. Step 1: non-inferiority Test (one-sided 97.5% CI) to show the investigational drug is as good as or not worse than the reference drug by a predefined margin (2,3).
         ii. Step 2: superiority Test (one-sided $\alpha = 0.025$) to show the investigational drug is superior to the reference drug. This test is performed only if non-inferiority in Step 1 is shown.
• Secondary comparison: superiority of the investigational drug 2X and 4X mg OD vs. reference drug Y mg.

b. Investigational drug vs. placebo

Primary analysis: dose-response via a regression analysis. A second-order regression analysis with the dose as predictor variable will be performed for the change from baseline in MSDBP at Visit 7 (Week 8) and endpoint to examine the relationship between the efficacy response and the dose. A test for lack-of-fit will be performed at significance level of 0.1.

Secondary analysis: pairwise comparison of investigational drug X, 2X and 4X mg OD vs. placebo.

The statistical test for each of the pairwise comparisons will be made at a two-sided 0.05 statistical significance level. Summary statistics for the changes from baseline of efficacy variables will be presented by treatment group and time point, as well as by treatment group, trial center, and visit; treatment group, age and visit; treatment group, sex and visit; and treatment group, race, and visit. Within-treatment analysis for all the efficacy variables will be performed by a paired t-test at the endpoint.

IV. OTHER STUDIES

PHASE IV MORTALITY AND MORBIDITY STUDIES

IV.1. Outline of a typical development plan

To evaluate the effect of antihypertensive therapy on cardiovascular events in hypertensive patients; the outcomes in subjects with hypertension who were treated with the new medication will be compared with the outcomes in those treated with conventional treatments. The study will be a double-blind, multicenter, randomized, parallel-group trial in subjects with essential hypertension (sitting blood pressure 160–200/95–115 mm Hg) or has been in some occasions a PROBE study (Prospective Open Blinded Endpoints).

IV.2. Long-term studies

IV.2.A. Objectives

Primary objectives
To compare the effect of either regimens in preventing cardiovascular complications either cardiac, cerebrovascular or a combination of endpoints which generally include cardiovascular death, acute MI, stroke and heart failure requiring hospitalization.

Secondary objectives
To compare the two regimens on some individual endpoints:
  a. hospitalization for angina, cardiac revascularization, heart failure, transient ischemic attack, accelerated or malignant hypertension, or renal failure in addition to the primary outcome;
  b. all-cause mortality;
  c. cancer.

IV.2.B. Primary endpoints

IV.2.C. Secondary endpoints

IV.2.D. Study design

Patients are randomized to once daily treatments under study-based antihypertensive treatment in a parallel-group for at least 4 years and until the calculated number of patients has a primary cardiovascular
event which has been validated (death, myocardial infarction, or stroke). The patients will be followed for the duration of period calculated to be necessary to see the predicted number of events. Patients are generally followed for a period of 4 to 6 years depending on the number of events required. They are followed at regular visits to obtain target blood pressures of less than 140/90 mm Hg. Central laboratory are required to standardize measurements such as ECG, echocardiogram or other specific measurements being used as endpoints.

As for other large morbidity and mortality trials, there are independent committees to adjudicate events and Data Monitoring safety Boards to guarantee the security and safety of the participants. The trial is generally run by a steering committee which determines a steering committee to organize the protocol.

Other medications
Usually, other drugs are permitted if clinically indicated. The choice of the drug will depend upon the drugs on study, in the sense that drugs of different class may be added. Any additional antihypertensive agent may be added as a step 3 medication (non-blinded).

IV.2.E. Planned sample
If the trial aims to document the effect of antihypertensive therapy on cardiovascular morbidity and mortality, patients with secondary hypertension, myocardial infarction or stroke within the previous 6 months, angina pectoris requiring treatment with β-blockers or calcium-antagonists, heart failure or left ventricular ejection fraction of 40% or less, or with a disorder that, in the treating physician's opinion, requires treatment with drugs of the same class of the tested drugs will be excluded.

The design of a phase IV study of morbidity and mortality requires the use of statisticians in the calculation of sample sizes based on the requirements determined by the investigators. These calculations take into account the primary endpoints, the potential duration of the study and the population under study to obtain predetermined objectives. It also assumes a percentage of non-compliance to the study, drop-outs to the trial and loss to follow-up during the trial. The population when comparing two active based treatments can be in the order of 15,000 patients up to 40,000 as was seen in the ALLHAT trail recently with a follow-up of 4 years (4).

IV.2.F. Study population

IV.2.G. Specific inclusion criteria

IV.2.H. Specific exclusion criteria

IV.2.I. Tools to assess endpoints
Efficacy assessment
The method of assessment will differ greatly depending upon the primary and secondary objectives of the trial. Trials have assessed the effect of test drugs on cardiovascular morbidity and mortality documenting left ventricular hypertrophy by means of electrocardiograms (5). In other trial, deaths were documented through the National Death Index; acute MI required two out of three of the following conditions:

a. symptoms compatible with acute MI (e.g., chest pain) lasting longer than 15 minutes;

b. electrocardiographic changes (new persistent ST-segment elevation or pathological Q waves in 2 contiguous leads); and

c. increased cardiac enzymes (more than twice the upper limit of normal).

A diagnosis of stroke required the presence of focal neurological deficit lasting longer than 24 hours. Imaging studies were not required to document a stroke. Any death thought to be compatible with
coronary heart disease (e.g., heart failure, sudden death) or cardiovascular disease was counted as a cardiovascular disease-related death (6).

An additional trial assessed the study outcomes at follow-up visits and reported to the clinical trials center. Hospitalized outcomes were primarily based on clinic investigator reports, and copies of death certificates and hospital discharge summaries were requested. In addition, searches for outcomes were accomplished through the Center for Medicare and Medicaid Services, the Department of Veterans Affairs, the National Death Index, and the Social Security Administration databases. A death was ascertained by clinic report or by match with the aforementioned databases plus a confirmatory death certificate. Medical reviewers from the clinical trials center verified the physician-assigned diagnoses of outcomes using death certificates and hospital discharge summaries. More detailed information was collected on a random subset of CHD and stroke events to validate the procedure of using physician diagnoses (4).

The frequency of the measure of endpoints depends on the objective of the trial. For instance, to assess cardiovascular endpoints, in one trial, participants were seen at least semi-annually for blood pressure measurements, treatment dispensing, and endpoint surveillance. On-site data verification was performed at least annually. An independent data and safety monitoring board met semi-annually to review accumulating data. Confidence intervals based on the Lan-DeMets version of the O'Brien-Fleming group sequential boundaries were used as guidelines for early termination. All analyses were performed independently of the sponsor. All study investigators and the study sponsors were blinded to all between-treatment comparisons until completion of endpoint data collection and review (6).

In another trial, patients were followed for at least 4 years with regular visits and increases in drug doses to reach a target blood pressure of less than 140/90 mm Hg. All screening, baseline, serial, yearly, and endpoint electrocardiograms were centrally assessed for signs of LVH and Minnesota coded at one reading center. Since combined ECG assessment of QRS voltage and duration enhances sensitivity for detection of LVH at acceptable levels of specificity, the product of QRS duration and Cornell voltage (with adjustment of 8 mm in women and a partition value of >2440 mm×ms) was used to recognize LVH. These composite ECG criteria have about 95% specificity in healthy people and 50% sensitivity in patients with LVH ascertained at necropsy or by echocardiography LVH (5).

**IV.2.J. Specific criteria for early withdrawal and discontinuation**

As described for phase II studies in II.2.J.

**IV.2.K. Data analysis method**

In one trial, time to event methods (Cox proportional hazards model and Kaplan Meier curves) were used to compare outcomes for participants randomly assigned to the investigational drug and the comparators. Analyses were by modified intent to treat (modified by the exclusion of 2 sites with data integrity concerns), unless otherwise specified, and were stratified by the choice of standard of care and geographic region in which the participant's clinical site was located. Analyses of primary and secondary events considered censoring due to losses to follow-up (e.g., participants for whom the primary event status was unknown on the closing date), non-cardiovascular disease-related deaths (as appropriate), and the closing date of the study. Losses were censored at the date the primary event status was last known (either the date provided by the site during the closeout process; or the date of the last follow-up visit). The proportional hazards assumption was tested by including an interaction term between the randomized treatment indicator and log-transformed follow-up time. Blood pressure changes from baseline were compared between the 2 treatment groups using the t test. All analyses were performed using SAS statistical software (Version 8.0, SAS Institute Inc, Cary, NC) (6).

In the second trial, analysis of all cardiovascular endpoints was by intention to treat; all randomized patients were included in their treatment group, and all available follow-up data were included from
randomization to the end of the study. The difference between treatment groups with respect to clinical events was assessed by a Cox regression model with degree of LVH (measured as a continuous variable) and the Framingham risk score defined by baseline characteristics as covariates. Treatment effects were measured by hazard ratios (relative risks) and 95% CIs by Cox regression models. The risk reduction for drug XXX against drug YYY was calculated as 100×(1–relative risk). Event rates over time are presented as Kaplan-Meier curves. Adjustment for blood pressure was derived from Cox regression models with blood pressures throughout the trial as time-varying covariates. Differences between groups in changes in ECG measures of LVH were analysed with the Wilcoxon rank-sum test, and the frequency of adverse experiences with Fisher's exact test (5).

V. REFERENCES


VI. EXAMPLES OF LANDMARK WELL DESIGNED TRIALS


APPENDIX 1 – Canadian Hypertension Recommendations

Recommended techniques for measuring blood pressure

1. Measurements should be taken with a sphygmomanometer that is known to be accurate. Although a mercury manometer may be preferable, a recently calibrated aneroid, or a validated and recently calibrated electronic device, can be used. Aneroid devices and mercury columns need to be clearly visible at eye level.

2. Choose a cuff with an appropriate bladder width matched to the size of the arm. The optimal bladder width equals the arm circumference/2.5, with an acceptable range of 80% to 100% of the arm circumference.

3. Place the cuff so that the lower edge is 3 cm above the elbow crease and the bladder is centred over the brachial artery. The patient should be resting comfortably for 5 min. in the seated position with back support. The arm should be bare and supported with the antecubital fossa at heart level because a lower position will result in an erroneously higher systolic blood pressure and diastolic blood pressure. There should be no talking, and the patient’s legs should not be crossed. At least two measurements should be taken in the same arm with the patient in the same position. Blood pressure also should be assessed after 2 min. of standing and at times when patients report symptoms suggestive of postural hypotension. Supine blood pressure measurements may also be helpful in assessing elderly and diabetic patients.

4. Increase the pressure rapidly to 30 mmHg above the level at which the radial pulse is extinguished (to exclude the possibility of a systolic auscultatory gap). Continue to auscultate at least 10 mmHg below phase V to exclude a diastolic auscultatory gap.

5. Place the bell or diaphragm of the stethoscope gently and steadily over the brachial artery.

6. Open the control valve so that the rate of deflation of the cuff is approximately 2 mmHg/heart beat. A cuff deflation rate of 2 mmHg/heart beat is necessary for accurate systolic and diastolic estimation.

7. Read the systolic level – the first appearance of a clear tapping sound (phase I Korotkoff) – and the diastolic level B the point at which the sounds disappear (phase V Korotkoff). Record the blood pressure to the closest 2 mmHg on the manometer (or 1 mmHg on electronic devices), as well as on the arm used, and note whether the patient was supine, sitting or standing. Record the patient’s heart rate. The seated blood pressure is used to determine and monitor treatment decisions. The standing blood pressure is used to examine for postural hypotension, if present, which may modify the treatment.

8. If Korotkoff sounds persist as the level approaches 0 mmHg, then the point of muffling of the sound is used (phase IV) to indicate the diastolic pressure.

9. In the case of arrhythmia, additional readings may be required to estimate the average systolic and diastolic pressure. Isolated extra beats should be ignored. Note the rhythm and pulse rate.

10. Leaving the cuff partially inflated for too long will fill the venous system and make the sounds difficult to hear. To avoid venous congestion, it is recommended that at least 1 min. should elapse between readings.

11. Blood pressure should be taken at least once in both arms, and if an arm has a consistently higher pressure, then that arm should be clearly noted and subsequently used for blood pressure measurement and interpretation.
Ambulatory blood pressure measurements will be made using the Spacelabs Model 90207 monitor.

On the days that the ABPM equipment will be applied, patients should arrive at approximately 8h00 AM to allow additional time for ABPM procedures such that dosing of medication occurs as close to 9:00 AM as possible.

The ABPM monitors will be programmed to measure blood pressure every 30 minutes throughout the day (0500 – 2300) and every 60 minutes at night (2300 – 0500). Patients will be advised not to move the arm during each blood pressure measurement and will also be given instructions concerning interruption of measurement in case of malfunction of the device or repositioning of the cuff if it slips.

For each of the two 24-hour ABPM monitoring sessions, the following procedures will be performed:
1. **Do not reuse batteries.** Always install four (4) fresh AA batteries prior to initializing the monitor.
2. Connect the cable from the modem to the monitor.
3. When attaching the monitor to the patient, first palpate the patient’s brachial artery (you can mark location with felt-tip pen) then apply an appropriate sized ambulatory BP cuff to patient’s non-dominant arm (e.g. if patient is right-handed, apply cuff to left arm).
4. Take up to five correlation readings. Attach T-tube to office sphygmomanometer, monitor and cuff. Allow cuff to inflate and listen to pressure. The monitor must be within 10 mm Hg of the pressure obtained by the column of mercury. If not, adjust cuff and try again.
5. Remove T-tube and attach cuff to monitor.
6. Manually trigger one or two ABPM readings to make sure the monitor is working properly. Give the patient a dose of study medication and then manually trigger another ABPM reading. The official “dose time” for the 24-hour recording will be the time shown on the ABPM clock as you pressed the blue start/stop button for the manual reading. Do not use your watch or a wall clock. Record the noted ABPM clock time on the CRF as the dose time. Dosing must occur at 8:00 AM plus or minus one hour.
7. Just prior to removal of the monitor, take a final manual reading using the start/stop button. The final manual reading should occur as close to 24 hours since prior dosing as recorded on the CRF. (The patient might need to repeat the ABPM if the monitoring is < 24 hours in duration, reference Appendix 11.3.3.)
8. After downloading of your patient’s data is complete, you may initialize the monitor for future use, making certain to first install fresh batteries. Make sure the monitor is turned off for storage.

The data collected will be evaluated to determine if it meets the criteria for a successful monitoring session.

**Operating instructions for ambulatory blood pressure monitoring**

Refer to the operating manual for a detailed explanation of ABPM operation.

**HANDLING OF ABPM DATA FOR ANALYSIS**

All data editing will be performed on blinded data with no further editing performed once patient treatment assignments are known.

**Screening Rules for Individual Readings**

The following screening rules will be used by the ABPM vendor to evaluate the validity of the individual readings from a patient’s monitor.

Screening Rules For Individual 20-Minute Interval Readings:
1. If the observed systolic blood pressure reading is either $<50$ mm Hg or $>250$ mm Hg, then the entire monitoring record will be considered invalid.

2. If the observed diastolic blood pressure reading is either $<20$ mm Hg or $>130$ mm Hg, then the entire monitoring record will be considered invalid.

3. If the calculated pulse pressure (i.e. SBP minus DBP) is either $<15$ mm Hg or $>150$ mm Hg, then the entire monitoring record will be considered invalid.

4. If the observed pulse rate reading is either $<20$ bpm or $>200$ bpm, then the entire monitoring record will be considered invalid.

An entire monitoring record refers to all readings and/or calculations, e.g. SBP, DBP, mean arterial pressure (MAP), pulse pressure (PP), and pulse rate (PR), for a particular 20-minute interval.

In addition, prior to statistical analysis of the ABPM results, the following screening rule will be used to further determine the validity of the individual SBP, DBP, and PR readings from a patient’s monitor.

5. For each observed reading, the six readings surrounding the observed reading (i.e. three before and three after) will be averaged together (the average will exclude the observed reading). If the observed reading differs by more than three standard deviations of the mean of the six surrounding readings and is outside of the range of values considered plausible for that particular measurement (see below) then the observed reading will be considered invalid. For observed readings that are numbered 1, 2, 3, 3rd to last, 2nd to last, or last, a wrap-around technique will be used to apply this screening rule.

   Plausible Range of Observed Readings:
   - SBP: mean ± 40 mm Hg
   - DBP: mean ± 20 mm Hg
   - PR: mean ± 20 bpm

**Calculation of hourly means**

Hourly means relative to both the dosing time and clock time will involve only valid 20-minute interval ABPM records. No imputation of missing or invalid readings/records will be performed. Additionally, one valid 20-minute interval reading per hour will be adequate for a hourly mean for each of the variables (e.g. SBP, DBP, MAP, PP, and PR).

**Hourly means relative to dose time**

Hour relative to dose time will be defined starting at dose time and incremented every 60 minutes. Valid readings collected at dosing and up to (but not including) one hour after dosing will be averaged to yield a 1-hour post-dose mean, valid readings collected at one hour post-dose and up to (but not including) two hours post-dose will be averaged to yield a 2-hour mean, etc. Hourly means will then be calculated for each of the hours post-dose that the monitor recorded measurements.

For example, if dosing was at 08:14 then records taken at 08:14 to 09:13 will be considered as within dose time hour 1, records taken at 09:14 to 10:13 will be considered as within dose time hour 2, etc., with hourly means relative to dose time calculated accordingly.

**Hourly means relative to clock time**

Clock time hour will be defined as given below with hourly means calculated for each clock time hour.

<table>
<thead>
<tr>
<th>Time Interval</th>
<th>Clock Time Hour</th>
</tr>
</thead>
<tbody>
<tr>
<td>24:00 – 00:59</td>
<td>0</td>
</tr>
<tr>
<td>01:00 – 01:59</td>
<td>1</td>
</tr>
<tr>
<td>02:00 – 02:59</td>
<td>2</td>
</tr>
<tr>
<td>03:00 – 03:59</td>
<td>3</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Time Interval</th>
<th>Clock Time Hour</th>
</tr>
</thead>
<tbody>
<tr>
<td>12:00 – 12:59</td>
<td>12</td>
</tr>
<tr>
<td>13:00 – 13:59</td>
<td>13</td>
</tr>
<tr>
<td>14:00 – 14:59</td>
<td>14</td>
</tr>
<tr>
<td>15:00 – 15:59</td>
<td>15</td>
</tr>
</tbody>
</table>
Readings of valid records taken at the end of a monitoring that are less than 24 hours since dosing and which would have the same clock time hour as the first clock time hour defined at the beginning of the dosing period will be used. All readings taken at or beyond 24-hours after dosing will not be used in the calculation of hourly means relative to clock time. For example, if dosing was at 08:14 then records taken the following morning from 08:00 to 08:13 will be used in the calculation of the hourly mean for clock time hour 8 while all records taken after 08:13 will not be used.

Criteria for a successful monitoring
The following rules will be used to evaluate whether the entire 24-hour interval of readings is unsuccessful and may need to be repeated.

Criteria for Successful Monitors:
The following types of monitors will not be considered successful:
1. those with more than a total of six non-consecutive hourly means missing during the 24-hour dosing period, or
2. those with more than three consecutive hourly means missing during the entire 24-hour dosing period.

Although there are potentially inherent differences in the hourly means relative to dose time and relative to clock time, as well as inherent differences due to the method of defining clock time hour, determination of whether or not a monitoring is “successful” and potentially needs to be repeated will be based on the ABPM vendor’s calculations.

ABPM Derived Endpoints
For all successful monitorings at baseline the ABPM vendor will calculate the 24-hour mean relative to clock time for diastolic blood pressure (DBP) to determine whether the patient qualifies for randomization into the active treatment phase.

Additionally, for all successful ABPMs the following endpoints as defined below will be derived:
1. Last 6-hour mean for systolic blood pressure (SBP), diastolic blood pressure (DBP), mean arterial pressure (MAP) and pulse rate (PR), based on the hourly means relative to dose time,
2. 24-hour mean for SBP, DBP, MAP and PR, based on the hourly means relative to dose time,
3. Morning (06:00-11:59) mean for SBP, DBP, MAP and PR, based on the hourly means relative to clock time,
4. Daytime (06:00-21:59) mean for SBP, DBP, MAP and PR, based on the hourly means relative to clock time,
5. Night time (22:00-05:59) mean for SBP, DBP, MAP and PR, based on the hourly means relative to clock time,
6. Systolic and diastolic load (i.e. the overall percentage of valid measurements for SBP above 140 mm Hg during the daytime and above 130 mm Hg during the night time, and the overall percentage of valid measurements for DBP above 90 mm Hg during the daytime and above 85 mm Hg during the night time, respectively).
ABPM response rate definitions are:
1) ABPM DBP "control" rate: 24-hour mean DBP < 80 mm Hg
2) ABPM DBP "response" rate: 24-hour mean DBP < 80 mm Hg or a reduction from baseline of ≥ 10 mm Hg
3) ABPM SBP "response" rate: 24-hour mean SBP < 130 mm Hg or a reduction from baseline of ≥ 10 mm Hg.

Graphical Presentation of Hourly Means
The hourly ABPM means (relative to dosing time) at baseline and at the end of the study will be averaged over all patients to get overall mean blood pressure profiles over the 24 hour post-dose period. These mean profiles will be graphically displayed for all treatment groups.

The mean profiles of the change from baseline in hourly means (relative to dose time) will also be graphically displayed for all treatment groups.
It is only intended to perform these analyses for the primary ABPM analysis dataset as specified in the STATISTICS section of the protocol.
Chapter 12. Lipid Lowering Agents

Robert Dufour, M.D.

Directeur associé
Unité de recherche en hyperlipidemies et athérosclerose
Institut de recherches cliniques de Montréal
120, avenue des Pins Ouest
Montréal, Québec H2W 1R7
CANADA
Tel 514 987-5758
Email robert.dufour@ircm.qc.ca
I. INTRODUCTORY REMARKS

Atherosclerotic cardiovascular disease is still the number one killer and the main cause of morbidity in the world. The link between high blood cholesterol, atherosclerosis and coronary heart disease (CHD) has been known for decades, but the scientific community has paid attention to this major risk factor only since powerful lipid lowering agents (LLA) were developed and commercialized about 15 years ago. The new statins (HMG Co A reductase inhibitors) can lower total and the most atherogenic LDL-cholesterol by up to 50-55% at maximum dosage. The beneficial effect of statins on plasma lipoproteins is reflected by a significant reduction in lesions progression and major cardiovascular events (by 25-40%) as demonstrated by many prospective primary and secondary prevention trials. The fibrates, another class of LLA that primarily decrease plasma triglycerides and increase the anti-atherogenic HDL-cholesterol, have also shown in angiographic and event trials a 25% reduction of major cardiovascular end points. More recently, the use of a statin to decrease LDL-C in combination with nicotinic acid to increase HDL-C in secondary prevention gave an even more spectacular 90% reduction of recurrent cardiovascular events after three years of treatment.

If the question of the benefit of treatment with LLA on cardiovascular disease is no more debatable, many issues regarding the use of LLA in clinical practice are still unresolved. For instance, in vitro and animal studies have documented various pleiotropic beneficial effects of statins on coagulation, fibrinolysis, oxidation and inflammation. Since these pathophysiological mechanisms are fundamental to atherosclerosis progression, it will be of great interest to assess whether the beneficial effect of statins is only related to lipid lowering or it is also mediated by other mechanisms. In patients with low lipid levels who are at risk of CHD because of intravascular inflammation or oxidation it remains unknown whether they would benefit from statins. Another fundamental question regards the optimal level of plasma cholesterol. It is still questioned whether the “lower is better” and whether there is a threshold effect to the lipid lowering benefit. Obviously many combination therapy trials will have to be conducted to resolve this issue.

Concerning the long term safety of LLA and recognition of adverse effects little is known. The longer these agents are used, the more we learn about subtle presentation of side effects. The neuromuscular symptoms associated with statin use or the fibrate induced fatigue are a few examples. These unrecognized side effects are also certainly involved in the problem of poor compliance to LLA. It is well documented, particularly in the elderly population, that after one year of treatment only 25% of the patients are still taking their statin. Even more worrisome is the possibility that a patient may suffer an unrecognized side effect that limits their quality of life. The issue of safety is crucial to permit the use of LLA in the paediatric population. Since statins provide effective prevention of atherosclerosis and CHD to patients with genetic diseases such as familial hypercholesterolemia in which cholesterol deposition in arteries starts in childhood, it is mandatory to establish the safety of statins at younger ages.

Some forms of morbid dyslipidemias such as hyperchylomicronemia and secondary pancreatitis are still untreated by drugs. Fortunately microsomal triglyceride transfer protein (MTP) inhibitors are in development and will need to be tested in these severe hypertriglyceridemic conditions. Other agents targeting enzymes and receptors of lipid metabolism such as cholesterol transfer protein (CETP) inhibitors, acyl transfer protein inhibitors, bile acid transporter inhibitors, LDL antioxidants etc. will also increase our arsenal of LLA in the next few years. These drugs will certainly open the way to combination therapies but may be also be the cause of lipid lowering drug interactions.

II. PHASE II STUDIES FOR REGISTRATION OF NEW DRUGS

II.1. Outline of a typical development plan
During this phase the lipid lowering effect of the new drug is compared to placebo in a population of moderately hyperlipidemic patients over a relatively short period of time (12-24 weeks). These controlled trials are initiated once earlier studies in normal volunteers have assessed the pharmacokinetics, optimal and
maximally tolerated dosages, dose schedule and interactions with food and other drugs. The Phase II trials are designed to evaluate the efficacy and short-term safety of the new agent. The assessment of efficacy of lipid lowering drugs is not free of ambiguity. LDL-C is an accepted surrogate for coronary heart disease, the ultimate aim of lipid lowering, and thus evidence that a drug lowers LDL-C by at least 15% is adequate for registration. However, for HDL-C raising and TG lowering drugs, not proven surrogates for CHD, it is likely that in addition to showing a benefit on lipoproteins, the demonstration of a benefit on clinical CHD outcomes will also be needed.

A multicenter, randomized, parallel group design with 2 to 4 groups receiving several drug doses, and one placebo group is generally used. Regulatory agencies require a minimum of two drug dosages and 12 weeks of active treatment for approval. A minimum of two dosages is necessary for observation of a dose versus effect relationship. The number of patients randomized is dependant upon the magnitude of the lipid lowering effect and the drop out rate expected. Patients with secondary dyslipidemia and a history of cardiovascular disease (secondary prevention) are excluded since the risk of placebo treatment for 3-4 months in this last group, is considered unethical. Because treatment with LLA is always additive to diet and lifestyle modifications, a pre-randomization run-in period of 4-8 weeks on diet and placebo is mandatory. Patients included in these short-term trials are usually allowed to enter subsequent long-term follow-up trials.

II.2. Short-term studies

Monotherapy trial for treatment of primary dyslipidemia

II.2.A. Objectives
   a. To determine the safety and efficacy as monotherapy for primary hypercholesterolemia.

II.2.B. Primary endpoints
   a. The percent change of LDL-cholesterol from baseline (randomization visit) to the end of the study.
   b. The incidence and prevalence of adverse events and clinical safety laboratory parameters abnormalities.

II.2.C. Secondary endpoints
   a. Percentage change in lipoprotein cholesterol and triglycerides (total, LDL-C, HDL-C, VLDL-C).
   b. Percentage change in apolipoproteins (apoA1, apoB, LDL apoB and Lp(a)).
   c. Percentage of patients meeting the NCEP (National Cholesterol Education Program) guidelines target lipid values at the end of the study.
   d. Percentage change in fat-soluble vitamin levels (vitamin E, A, D, ß-carotene) and INR as a functional measure of vitamin K status (for drugs affecting lipid absorption).
   e. Quality of life assessed by the SF-12 Health Survey.

II.2.D. Study design
A multicenter, randomized, double-blind, parallel group, placebo controlled design is generally used. Study begins with a screening phase using patient’s file, questionnaire, physical examination and blood tests to identify those meeting the clinical and lipids inclusion/exclusion criteria. Potential candidates then enter a 4-8 weeks lead-in, stabilization period, during which they will receive single-blind placebo medication, lifestyle and diet counselling according to NCEP guidelines. A minimum of six weeks without LLA is needed before randomization. The lead-in period is also used to monitor drug compliance. Those not taking 75-80% of their medication are usually excluded. At this point patients fulfilling the inclusion/exclusion criteria are randomized to placebo or one of two active treatment arms (with one or up to four different doses of study medication). Sometimes a short titration phase may be used within the active treatment arms. During the treatment period lasting generally 12 to 24 weeks, the dosage remains stable until the end of the study. No other lipid lowering or lipid affecting agents are allowed during the
study, and an effort is made to keep the concomitant medications stable. At the end of the study, each participant is usually invited to take part in a long term extension trial with a predetermined randomization to an active drug treatment regimen.

II.2.E. Planned sample
Assuming a dropout rate of 15 to 25% over a 24 weeks period, approximately 150 patients per group are needed to detect, with 90% power, a significant difference of 15% \( (p \leq 0.05) \) between each treatment arm and placebo.

II.2.F. Study population
Adults (over 18 years) with primary hypercholesterolemia.

II.2.G. Specific inclusion criteria
a. Adults over 18 years of age at screening.

b. Females must not be pregnant or lactating. Females of childbearing potential, must use a medically acceptable form of contraception at least four weeks before and until four weeks after the end of the study.

c. Documented history of hypercholesterolemia with LDL-cholesterol (a mean of the 2 values obtained at the screening and randomization visits) \( \geq 3.4 \text{ mmol/L} \) and \( \leq 6.2 \text{ mmol/L} \) and TG \( \leq 3.5 \text{ mmol/L} \).

d. Have the ability to comply with the NCEP Diet.

II.2.H. Specific exclusion criteria
a. History of atherosclerotic vascular disease including CABG (Coronary Artery Bypass Graft), PTCR (Percutaneous Transluminal Coronary Revascularisation), clinical or symptomatic angina pectoris, myocardial infarction, stroke, or peripheral vascular disease.

b. Uncontrolled primary hypothyroidism (as defined by TSH \( > 1.5 \) times the upper limit of normal at the screening visit), nephritic syndrome, or other possible causes of secondary dyslipidemia.

c. Hemoglobin \( \text{AIC} > 8.0\% \).

d. History of cancer in the past five years (excluding basal cell carcinoma).

e. Uncontrolled hypertension as defined by a systolic BP \( \geq 160 \text{ mmHg} \) or diastolic BP \( \geq 100 \text{ mmHg} \).

f. Chronic renal failure or serum creatinine \( > 1.7 \) times the upper limit of normal at the screening visit.

g. Unexplained serum CK \( > 3 \) times the upper limit of normal at the screening visit.

h. Active hepatitis or cholestasis or ALT \( > 2 \) times the upper limit of normal at the screening visit.

i. History of drug or alcohol abuse within the last year (more than 21 alcoholic beverages/week).

j. Chronic diarrhoea or malabsorption.

k. Subjects taking one of the following medications and unable to maintain a stable dose at least four weeks prior to the screening visit and for the duration of the study: tamoxifen, raloxifene, estrogen and or progestins, thiazide diuretics, isotretinoin, \( \beta \)-blockers, thyroid hormones, androgens, fiber supplements, protease inhibitors.

l. Subjects currently taking and unable to discontinue prior within eight weeks prior to randomization and throughout the study: Any lipid lowering agent, cyclosporine, orlistat, systemic corticosteroid, alpha-glucosidase inhibitors.

m. Drug compliance lower than 80% of the expected tablet count during the last two weeks of the lead-in period.

n. Have received any investigational drug within four weeks prior to the screening visit.

II.2.I. Tools for assessing primary endpoints
Blood tests.
II.2.J. Specific criteria for early withdrawal and discontinuation

The criteria for discontinuation from the study are pre-specified:

- Failure to meet randomization criteria;
- Protocol non-compliance;
- Adverse events;
- Investigator judgment;
- Patient withdraws consent;
- Pregnancy;
- ALT $\geq 3$ times the upper limit of normal;
- CK $\geq 10$ times the upper limit of normal or CK $\geq 5$ times the upper limit of normal with clinical signs of myopathy; on two consecutive occasions at least one week apart.

II.2.K. Data analysis method

Baseline homogeneity of the variables is examined by a one way ANOVA. The analysis of efficacy is done according to the intention-to-treat principle. All statistical tests are two sided and p values $\leq 0.05$ are considered statistically significant. The only multiplicity adjustments are the Bonferroni adjustments for the primary hypothesis tests, one for each of the active treatment arm vs. the placebo.

II.3. Long-term studies

Following completion of the short term, double blind, placebo controlled, efficacy and safety study, subjects are often offered entry into an open-label, long term (usually 12 months), extension study. The purpose of the trial being to gather data on long-term safety and tolerability of the study drug. In monotherapy long-term trials, during the first few weeks, the lipid lowering drug is usually titrated up (when it is possible), for subjects not attaining at the end of the short-term trial, the LDL-cholesterol goal dictated by NCEP guidelines. In combination trials, it may be the combined agent that is titrated up. After the titration period, usually patients are followed less frequently (every 3 months). Still, if necessary to reach therapeutic goal, other lipid lowering agents may be added during long-term follow-up. To demonstrate safety, the ICH guidelines generally ask that 600 patients be treated for 6 months with the new agent and 100 patients studied for one year. These extension studies offer also the possibility to test the persistence of the lipid lowering effect.

III. PHASE III STUDIES FOR REGISTRATION OF NEW DRUGS

III.1. Outline of a typical development plan

The phase III development plan of a lipid lowering drug must include a double-blind, randomised, placebo-controlled parallel group study which replicate the short-term phase II study, but generally uses only the optimal dosage and recruits a larger number of patients. The study is performed in duplicate usually in different populations; one in North America and one in Europe for example. The U.S. Food and Drug Administration (FDA) requires that phase III studies need to recruit a significant number of women and non-Caucasian participants. The main objective of phase III studies is to test the reproducibility and the expandability of the phase II results to the general population.

The number and the nature of the other trials in the development plan will be dictated by the intended use and indication of the drug. For instance, if a strong and well tolerated hypocholesterolemic agent with a new mechanism of action is intended to be used in replacement of statins to decrease LDL-cholesterol lowering, the phase III plan will include:

- Trials comparing the most powerful statins on the market with the new agent, in moderately and severely hypercholesterolemic subjects.
b. The same comparison in special populations such as patients with Heterozygous Familial Hypercholesterolemia (FH) adult or children, Homozygous FH, renal failure, sitosterolemia or aortic stenosis.

c. Combination trials with fibrates in patients with mixed dyslipidemia.

d. Long-term outcome trials (recurrent myocardial infarction or cardiovascular death) in secondary prevention.

e. Trials measuring the benefit of the drug on surrogate markers of atherosclerosis such as endothelial dysfunction, intima-media thickness or coronary plaques evaluated by intravascular ultrasound (IVUS) or angiography.

Before testing the new drug in children, elderly people or patients with renal failure it is necessary to demonstrate the pharmacokinetic and pharmacodynamic equivalence in a few subjects (8-10). A trial performed in a paediatric population with Familial Hypercholesterolemia is outlined below as an example of a phase III study.

### III.2. Short-term monotherapy studies

**Monotherapy trial in children with Heterozygous Familial Hypercholesterolemia**

#### III.2.A. Objectives
To evaluate the efficacy and safety as a monotherapy agent in adolescents with heterozygous familial hypercholesterolemia (HeFH).

#### III.2.B. Primary endpoints
- Percentage of change of LDL-cholesterol from baseline to six months.
- Incidence of adverse events, ECG and clinical laboratory safety parameters.

#### III.2.C. Secondary endpoints
- Specific apolipoproteins, lipoprotein cholesterol and triglycerides.
- Serum steroid hormone levels
- Serum fat-soluble vitamin levels
- Serum vitamin B-12 and serum/red blood cell folate levels
- INR as a functional measure of vitamin K status
- Linear growth

#### III.2.D. Study design
A multi-center, double-blind, placebo-controlled, parallel group, randomized study to evaluate the safety and efficacy in monotherapy. The study consists of a five to eight week drug washout, diet stabilization lead-in period, and a twenty-six (26) weeks double-blind treatment period. At the end of the lead-in period, patients meeting the randomization criteria will be randomized to one of the two arms: Active drug at the optimal dose or placebo. An unbalanced randomization can be used such that the ratio of patients on active drug/placebo will be 2/1.

#### III.2.E. Planned sample
Sample size will be similar to what has been described for phase II monotherapy trials. But since the study includes adolescents, it is wise to recruit a larger number of patients to counteract a significant Lost to follow-up. With the same purpose, the number of visits should be kept minimal.

#### III.2.F. Study population
Male and female adolescents between 10 and 17 years of age with HeFH.
III.2.G. Specific inclusion criteria
a. HeFH documented by LDL-cholesterol levels consistent with FH (LDL-C > 4.2 mmol/L) or documentation of LDL-cholesterol receptor gene mutation AND a positive family history of atherosclerosis at or before 50 years of age in males, and before 60 years of age in females, OR documented family history of hyperlipidemia with LDL cholesterol levels above 95th percentile for age and sex before treatment.
b. The participant and the parent/legal guardian must provide informed consent prior to the subject undergoing any study-specific screening procedures.

III.2.H. Specific exclusion criteria
a. As described in the monotherapy phase II trial for treatment of primary dyslipidemia.
b. History of homozygous familial hypercholesterolemia.

III.2.I. Tools for assessing primary endpoints
Blood tests.

III.2.J. Specific criteria for early withdrawal and discontinuation
As described in the monotherapy phase II trial for treatment of primary dyslipidemia.

III.2.K. Data analysis method
As described in the monotherapy phase II trial for treatment of primary dyslipidemia.

Long-term monotherapy studies
As described in the monotherapy phase II trial for treatment of primary dyslipidemia.

IV. PHASE III STUDIES FOR REGISTRATION OF NEW DRUGS: ADJUNCTIVE THERAPY
INDICATIONS

IV.1. Outline of a typical development plan

At the present time, the vast majority of patients treated for dyslipidemia receive only one lipid agent (monotherapy). There are numerous reasons for that: First, the NCEP has designated LDL-C reduction as the primary target of treatment to prevent atherosclerosis and statins used in monotherapy can normalize LDL-C for most of the patients. Second, most of the long-term event trials that have proven the benefit of lipid lowering to prevent CVD have been done with monotherapy. Third, combinations of certain lipid lowering agents (i.e gemfibrozil+ statin) have been proven hazardous in the past, with cases of rhabdomyolysis and acute renal failure. Fourth, in patients with mixed dyslipidemia (elevated LDL-C and TG), the hypertriglyceridemia is often secondary to overweight, a poor diet, lack of exercise or alcohol intake, and can be corrected by lifestyle modifications. Fifth, physicians are reluctant to use two different drugs in asymptomatic patients to reduce, at significant costs, a single risk factor of CVD, particularly in primary prevention.

Nevertheless, some clinical situations require combinations of lipid lowering agents to effectively treat morbid forms of dyslipidemias. A combination of a full dose statin and ezetimibe is often necessary to normalize LDL-C in HeFH patients. A statin associated with nicotinic acid can effectively take care of high LDL-C and low HDL-C, a form of dyslipidemia often encountered in survivors of myocardial infarction. High risk patients with severe mixed dyslipidemia may need a combination of a fibrate and a statin for optimal therapy.
New drugs in development may have relatively modest cholesterol lowering effect but a complementary anti-atherosclerotic mode of action on oxidation, inflammation or thrombosis. The phase III development design for such agents intended to be used as adjunctive therapy must minimally include:
   a. Two large scale, short-term, parallel group, trials in different populations comparing the effect of the optimal dose of the new drug used in combination with a statin vs the statin alone on lipid levels and surrogate markers of atherosclerosis, inflammation and oxidation.
   b. Another short-term trial using the new drug in combination with a fibrate compared with the fibrate alone in patients with mixed dyslipidemia.
   c. A long-term trial comparing in secondary prevention, the same treatment arms over cardiovascular endpoints.
   d. A long-term trial testing the effect of the combination with a statin on carotid atherosclerosis progression evaluated by measurement of intima-media thickness (the only method approved by FDA).

Some studies on LLA intended to be utilised in combination therapy have used a cross-over design where each patient is treated successively with each drug and then with the combination. In such study design each patient serves as his own control, and each of the three treatment periods has to be long enough (2-3 months) to prevent a carry-over effect of the previous treatment.

IV.2. Short-term adjunctive therapy studies

Combination-therapy trial in subjects with primary hypercholesterolemia at high-risk of cardiovascular disease not controlled by a starting dose of a statin alone

IV.2.A. Objectives
To evaluate the efficacy and the safety of the investigational drug administered in combination with a statin in subjects with primary hypercholesterolemia and at high risk of cardiovascular disease, who do not reach the target value of LDL-C (2.5 mmol/L) on starting dose of the statin alone.

IV.2.B. Primary endpoint
   a. The proportion of subjects achieving target LDL-C (according to NCEP ATP-III guidelines) after 12 weeks of treatment (the aim is to show that a greater number of patients will reach the LDL-C target by combining the investigational agent to the statin rather than by increasing the dosage of the statin).

IV.2.C. Secondary endpoints
   b. The proportion of subjects achieving target LDL-C at weeks 2, 4 and 8.
   d. Change in quality of life, evaluated by the SF-36 questionnaire, after 12 weeks of treatment.
   e. Adverse events and laboratory abnormalities occurring during the active treatment phase.

IV.2.D. Study design
A multi-center, randomized, double-blind, parallel group study.

Subjects with primary hypercholesterolemia and at high risk of cardiovascular event because of CHD history or diabetes mellitus or presence of other risk factors (absolute risk greater than 20% at 10 years according to Framingham tables) will be recruited. After a four weeks wash-out period of any lipid lowering agent (eight weeks for fibrates and one year for probucol), candidates will receive a starting dose of statin for another four weeks (open-label). At the end of the open-label period, they will be randomized to continue on statin therapy and placebo or to the combination of statin and the investigational drug at a
fixed dose for 12 weeks. In each group, the statin dosage is doubled after 4 and 8 weeks of treatment if LDL-C remains greater than 2.5 mmol/L. Diet therapy will be followed for the duration of the study.

**IV.2.E. Planned sample**
To show a 15% difference in the percentage of subjects achieving the target LDL-C level, with a power of 90% and a significance level of 0.05; a total sample of approximately 500 participants (250 per treatment arm) is needed.

**IV.2.F. Study population**
Men and women with primary hypercholesterolemia as well as with cardiovascular disease or well controlled diabetes mellitus or a cardiovascular risk greater than 20% at 10 years according to Framingham tables (NCEP ATP-III)

**IV.2.G. Specific inclusion criteria**

- **a.** Age 30 to 70 years.
- **b.** Primary hypercholesterolemia with an LDL-C ≥ 2.5 mmol/L and TG ≤ 3.5 mmol/L after the four weeks run-in period of open label statin therapy.
- **c.** Documented history of cardiovascular event (acute coronary syndrome; coronary, cerebral or peripheral revascularization; stroke or transient ischemic attack).
  
  \[ \text{AND/OR} \]
  
  Well controlled and stable diabetes mellitus (HbA1C ≤ 7.0% at baseline and no change in treatment later than 3 months before screening).

  \[ \text{AND/OR} \]
  
  A risk of CHD greater than 20% at 10 years, according to the modified Framingham tables of the NCEP ATP-III.

**IV.2.H. Specific exclusion criteria**

- **a.** Medical condition likely to limit life span to less than one year
- **b.** Women of child bearing potential who do not agree to practice an effective method of birth control until one month following study completion
- **c.** Lactating women
- **d.** Postmenopausal women on hormone replacement therapy
- **e.** Patients with homozygous familial hypercholesterolemia
- **f.** Uncontrolled endocrine or metabolic disease known to influence serum lipids or lipoproteins (thyroid hormone substitution must be stable for at least 3 months prior to screening)
- **g.** Chronic renal insufficiency or nephrotic syndrome
- **h.** Active or chronic hepatobiliary or hepatic disease
- **i.** Treatment with agents with known interactions with the statin or the investigational drug used in the study
- **j.** Any other lipid-altering agents (drug or food supplement) administered during the study

**IV.2.I. Tools for assessing primary endpoints**

Blood tests and questionnaire about quality of life (SF-36 Acute, Chapter 25, Appendix B).

**IV.2.J. Specific criteria for early withdrawal and discontinuation**

Threatening of health or well-being of a participant by their continuation in the study, in the opinion of the investigator.

**IV.2.K. Data analysis method**

The analysis of the primary efficacy variable is based on the intention-to-treat and performed using chi-square test. Analysis at each time point (4, 8 and 12 weeks) is provided.
**IV.3. Long-term adjunctive therapy studies**

Combination-therapy trial in subjects with mixed dyslipidemia

**IV.3.A. Objectives**
To demonstrate that the fixed combination of a new triglyceride lowering agent (TLA) with a cholesterol lowering agent (CLA), titrated to the NCEP ATP-III guidelines, can slow the progression of atherosclerosis when compared to the cholesterol lowering agent alone, in subjects with mixed dyslipidemia after 24 months of treatment.

**IV.3.B. Primary endpoint**
- a. Annualized rate of change in intima-media thickness (IMT).

**IV.3.C. Secondary endpoints:**
- b. Annualized rate of change in IMT of each anatomical site (internal carotid, bifurcation and common carotid).
- c. Change in lipid parameters and biochemical markers of vascular inflammation and insulin resistance.
- d. Cardiovascular events (combination of coronary death, non-fatal myocardial infarct, stroke and revascularization procedures).
- e. Adverse events and abnormalities of clinical laboratory parameters.

**IV.3.D. Study design**
A multi-center, double-blind, randomized, parallel group, carotid ultrasound study of the fixed combination compared to the cholesterol lowering agent alone administered to subjects with mixed dyslipidemia. Potential subjects will undergo a 4-week lead-in period when they will receive diet and lifestyle counselling only (no LLA), as described in the NCEP ATP-III clinical guidelines or equivalent to establish screening lipid levels. This is especially important in patients with mixed dyslipidemia since diet and lifestyle has great impact on triglyceridemia. At the end of the lead-in period, eligibility for entry into the study is determined. All eligible subjects will enter a run-in period of treatment with the cholesterol lowering agent alone. The dosage will be titrated-up every four weeks to a target LDL-C level as per NCEP ATP-III guidelines. Once the target LDL-C is reached, subjects will then be randomized to continue on the CLA alone or to the combination with the TLA at fixed dose. Carotid ultrasonography will be performed at baseline and every six months for 24 months of treatment. Clinical safety and/or lipid efficacy assessments will be performed at each visit.

**IV.3.E. Planned sample**
It is estimated from similar studies that the standard deviation of the annualized rate of IMT change is 0.06 mm/yr. With a sample size of 340 subjects in each treatment group and a two-sided alpha level of 0.05, the study will have 90% power to detect a difference of 0.015 mm/yr. With a dropout rate of 20%, 850 subjects will be randomized.

**IV.3.F. Study Population**
Men and women with mixed dyslipidemia persisting after diet and lifestyle modifications. Patients with dysbetalipoproteinemia (Type III) should be excluded because they are particularly responsive to TLA.

**IV.3.G. Specific inclusion criteria**
- a. Age 30 to 65 years
- b. Mixed dyslipidemia defined by LDL-C ≥ 4.2 mmol/L and ≤ 6.2 mmol/L AND TG ≥ 2.3 mmol/L and ≤ 6.5 mmol/L.
IV.3.H. Specific exclusion criteria
   a. Women who are pregnant or lactating.
   b. Patients with dysbetalipoproteinemia (Type III)
   c. Patients with symptomatic cardiac or cerebrovascular disease (within the last 3 months prior to the screening visit)
   d. Patients who do not have a satisfactory baseline carotid ultrasound.
   e. Patients requiring specific lipid lowering therapy different from the CLA used in the study.
   f. Subjects taking any drugs known to interact with the medications used in the study.
   g. Subjects with uncontrolled hypertension (>140/90 mmHg), diabetes or hypothyroidism.
   h. Subjects with history of acute pancreatitis, malignancy, hepatobiliary disease (active), nephrotic syndrome, chronic renal failure or malabsorption.
   i. Subjects with alcohol consumption of more than 14 alcoholic beverages per week.
   j. Subjects with severe obesity or body mass index > 35 kg/m2.
   k. Subjects with poor compliance during the run-in period (< 80%) as assessed by tablet count.

IV.3.I. Tool for assessing primary endpoints
Quantitative ultrasonography.

IV.3.J. Specific criteria for early withdrawal and discontinuation
Finding of severe atherosclerotic carotid stenosis requiring surgery or adverse event requiring discontinuation of treatment.

IV.3.K. Data analysis method
The analysis of safety and efficacy data is done on the intention-to-treat population. But to be included in the efficacy analysis a participant needs to have at least both a valid baseline and a valid follow-up carotid IMT evaluation performed at least one year post baseline. The primary endpoint comparison between the two treatment groups needs to include analysis of time by treatment interaction.

V. OTHER STUDIES (SPECIAL INDICATIONS)

V.1. Hyperlipidemia in the elderly
Since the benefit of treating hyperlipidemia to prevent cardiovascular events in elderly men and women (70-80 years) at high cardiovascular risk has been demonstrated in large scale randomized trials; the population involved in new drug trials should include a significant proportion of older patients. The number of drugs susceptible to be prescribed unilaterally to older people for secondary prevention of CVD is growing every year (aspirin, ACE inhibitors, beta-blockers, clopidogrel, statins, etc.). As a consequence, pharmacokinetic and interaction studies should be performed in elderly subjects. New drug phase III trials should not be restrictive in terms of concomitant medication in order to represent a “real life” situation and obtain valuable information on the safety of the new agent in this segment of the population. It may also be judicious in the future to include systematically in every long term trial involving subjects older than 70 years an evaluation of cognitive functions in order to detect subtle beneficial or deleterious effects.

V.2. Hyperlipidemia in dialysis and renal transplant patients
Total cholesterol level is inversely associated with survival in dialysis patients, a group at high risk of cardiovascular disease. The nephrotic syndrome is associated with high plasma levels of total and LDL-C, thus can promote atherogenesis and the risk of CHD. Chronic renal failure is characterized by hypertriglyceridemia, and decreased levels of HDL-C in approximately 30% of patients. Patients with kidney transplantation exhibit variable patterns of hyperlipidemia; some patients are purely hypertriglyceridemic while others present with hypercholesterolemia or mixed dyslipidemia. Unfortunately, the risk of myotoxicity and of interaction between cyclosporine and many of the currently available lipid lowering drugs (statins or fibrates), and the lack of large scale trials, has hampered their
utilization in this subgroup of patients. Thus, new drug development projects should include phase III short and long term studies to examine the efficacy and safety to reduce hyperlipidemia and CVD in patients with renal disease.

**VI. EXAMPLES OF LANDMARK WELL DESIGNED TRIALS**

**Monotherapy, short-term trials**

**Monotherapy, long-term trials**

**Adjunctive therapy trials**

**Clinical events trials**
Chapter 13. Anti-atherosclerotic Drugs

Jean-Claude Tardif, M.D. ¹
Therese M. Heinonen, D.V.M. ²
Marie-Claude Guertin, Ph.D. ²

¹ Research Center
Montreal Heart Institute
5000 Belanger Street
Montreal, Quebec  H1T 1C8
CANADA
Tel  514-376-3330 x 3612
Fax:  514-593-2500
Email  jean-claude.tardif@icm-mhi.org

² Research Center
Montreal Heart Institute
5000 Belanger Street
Montreal, Quebec  H1T 1C8
CANADA
I. INTRODUCTORY REMARKS

At this moment, atherosclerotic disease places a great burden on society. Not only is atherosclerosis currently one of the leading causes of death around the world, but it also has a major impact on an individual’s quality of life as a result of chronic pain, activity restriction, unemployment and disability. In 2002 in the United States alone, it was estimated that over 62 million individuals had one or more type of cardiovascular disease and that the direct and indirect cost of atherosclerotic coronary artery disease (CAD) was over 110 billion dollars.

A variety of risk factors for the development of atherosclerosis have been identified. These risk factors include smoking, high blood pressure, high cholesterol, diabetes, obesity, male gender, genetic predisposition, and physical inactivity. Lifestyle modifications and pharmacological therapy, including cholesterol-lowering medication, have become the standard of care provided to patients with coronary artery disease. While these interventions have been effective, they have failed to keep pace with the expanding “at risk” population. In particular, the increased proportion of elderly individuals and the rise in prevalence of risk factors such as metabolic syndrome, diabetes, and obesity are cause for concern. It is important to recognize that cardiovascular disease not only affects the elderly but is also a leading cause of premature death in individuals under the age of 75. Furthermore, hospitalization rates by age groups indicate that acute myocardial infarction (MI) and ischemic heart disease become important diseases by the fourth and fifth decades of life. As a result of this trend, there has been an increase in the number of deaths due to cardiovascular disease worldwide with the expectation that this trend will continue for the next fifteen years.

A complex series of events leads to the formation of atherosclerotic plaques in human arteries. Atherosclerosis is now understood to be a systemic chronic inflammatory disease characterized by the excess accumulation of lipid-laden macrophages (foam cells) within the arterial wall. The steps involved in the development of atherosclerotic lesions include endothelial expression of adhesion molecules, release of cytokines and chemokines, involvement of reactive oxygen species, macrophage accumulation in the arterial wall, and incorporation of oxidized LDL. The resulting atheroma is comprised of cholesterol, inflammatory cells, and matrix. Atherosclerotic lesions are heterogeneous and the clinical manifestation of atherosclerosis is dependent on the lesion composition and affected vascular bed, as well as other factors. When the atherosclerotic plaque ruptures the clinical presentations include MI, unstable angina, stroke, transient ischemic attack (TIA), and cardiac death.

The complex pathophysiology of atherosclerotic disease highlights the fact that many processes contribute to lesion development, and suggests that cholesterol modification is not the only mechanism by which the condition may be positively affected. It is well accepted that high serum levels of cholesterol, and particularly low-density lipoprotein cholesterol (LDL-C), increase the risk for CAD. The use of HMG-CoA reductase inhibitors (statins) became widespread as data from numerous clinical trials supported their safe and efficacious use in a broad population base. With a reduction in combined cardiovascular morbidity and mortality of 30 – 35% resulting from their use, statin drugs have had a substantial positive benefit on health care. However, the majority of clinical events are not prevented by statins and the persistently high level of cardiovascular disease is far from satisfying. Attention has therefore been focused on non-lipid risk factors of CAD including some of the key systems involved in the formation and progression of atherosclerotic plaque. As a result of the ubiquitous and chronic nature of atherosclerosis, as well as the persistently high incidence of cardiovascular morbidity and mortality, research efforts in this area remain vital. In recent years these efforts have changed our understanding of atherosclerosis significantly. A process which was once thought of as a static accumulation of fat in the arteries is now realized to be a very dynamic process in which numerous environmental, genetic, and individual factors contribute. Research which at one time focused almost exclusively on cholesterol metabolism has now branched out to provide further insight into the role of
other processes such as inflammatory contributors. A number of chemokines and cytokines involved in the inflammatory process have now been identified, and potential markers of disease progression and regression are under evaluation. Additionally, our ability to understand the process of atherosclerosis and the causative factors has gained sophistication with the experimental and clinical employment of various high-technology tools. These tools include the use of genomics, proteomics, the identification of soluble plasma markers, as well as imaging technologies such as quantitative coronary angiography, intravascular coronary and peripheral ultrasound, B-mode carotid ultrasound, positron emission tomography, computed tomography, and magnetic resonance imaging.

The development of atherosclerotic plaque changes both vascular structure and function. The endothelial lining of the blood vessels is important in the regulation of vascular tone and the accommodation toward greater blood supply in circumstances of greater demand. Modification of endothelial function is one of the earliest clinically demonstrated changes within the vasculature during the course of atherosclerotic progression. Subsequent lesion progression with the accumulation of cholesterol and matrix leads to clinically detectable changes in vascular anatomy and further functional changes. The vascular wall becomes thickened with accumulating debris, and the lumen eventually narrows despite remodelling changes to accommodate expanding plaque volume. The convergence of these factors, along with the inflammatory components, results in supply/demand mismatch as well as plaque instability and rupture – all accumulating in clinical disease.

While progress has been made in the field of atherosclerosis, there remain many challenges in developing an agent which will work by a non-lipid anti-atherosclerotic mechanism. There are currently no regulatory guidelines for the development of such drugs and no therapies approved for the treatment of atherosclerotic coronary disease that work by non-lipid mechanisms. Additionally, despite the recent generation of data identifying potential biomarkers such as C-reactive protein, clear and independent links between such markers and coronary artery disease still need to be strengthened in order to maximize their utility in drug development programs. Other developmental hurdles include the chronic nature of CAD, trial duration, ethical considerations, concomitant medications, and plaque heterogeneity. Since atherosclerotic plaques develop slowly over decades, in most instances, clinically meaningful changes using even the most sensitive methods have required relatively long-term therapy of 12-18 months. Due to the effectiveness of current therapy, including statins, at decreasing cardiovascular risk in a wide patient population, true placebo-controlled trials in patients with cardiovascular risk have become unethical. Also due to the demonstrated effectiveness of current therapy, any potential new therapy is expected to be evaluated as an adjunct to standard care. Standard care may involve several medications for lipid-lowering, as well as hypertension and glucose control. Care must therefore be given to potential interactions between new agents and these concomitant medications. While the development of new antiatherosclerotic agents that work by non-lipid mechanisms will continue to present a developmental challenge, the new tools available, and the diligence of those undertaking this challenge is expected to make significant inroads in our ability to successfully identify the next wave of effective therapy.

II. PHASE II STUDIES FOR REGISTRATION OF NEW ANTI-ATHEROSCLEROTIC DRUGS

II.1. Outline of a typical development plan

Although atherosclerosis is a systemic condition, which leads to stroke and other serious peripheral circulatory disorders, the development of anti-atherosclerotic agents typically focuses on coronary disease indications. Standard exploratory studies conducted during this stage of development include those designed to assess the maximally tolerated dosages over a short to moderate treatment durations, pharmacokinetics and drug interactions. The length of treatment should be sufficient to obtain the safety and efficacy information needed to determine the feasibility of the long-term evaluations needed in Phase III. Additionally, since these agents work by a non-lipid mechanism, it is important for the early clinical
trials to explore the utility of other potential biomarkers. Two types of biomarkers may be included in these trials. The first type of biomarker is one which will provide pharmacodynamic information on the biochemical system targeted. For example, an agent which activates enzyme X should be evaluated at this stage for the degree of activation conferred across an acceptable dose range, if possible. The second type of biomarker is one which will provide a reasonable indication of clinical efficacy (ie. vascular improvement). The selection of potential biomarkers for evaluation should be based on a good understanding of the agent’s mechanism of action, as well as the pathophysiology of atherosclerotic disease. The appropriate set of biomarkers for each evaluated agent may be quite different depending on the mechanism and anticipated response. Therefore the development plan for each agent may also be unique. Additionally, since lipid-lowering therapy has clearly demonstrated effectiveness and safety, it is important to assess at an early stage the influence of the study drug on current therapy through the use of well designed drug interaction studies. The results of these carefully designed early studies are critical for appropriate design of pivotal studies, which may be significant in terms of size, duration, and resource needs.

There are several different approaches employed in Phase II to obtain clinical efficacy. One approach is to evaluate a population with a clinical indication for cardiac catheterization in order to enable the use of sensitive but invasive imaging tools such as intravascular ultrasound (IVUS) to detect drug-induced changes. Such an approach has the advantages of directly assessing the coronary circulation, minimizing sample size and detecting small but potentially important changes in plaque burden. However, the obvious disadvantages are the invasiveness of the assessment and the limitations regarding patient selection. Another approach is to conduct initial evaluations in individuals with stable moderate atherosclerotic disease using non-invasive tools such as B-mode ultrasound in peripheral vessels (carotid and brachial). Such an approach has the advantage of non-invasive assessments in a widely available patient population, but relies on correlations between peripheral changes and coronary disease. Additionally, this approach may require a larger sample size to detect drug effect changes. Often, both of these approaches are included in the Phase II development plan. The typical pivotal trial will utilize a double-blind, multi-center, placebo-controlled, randomized, adjunctive therapy, parallel group design. At least 2 or 3 dose levels should be explored, preferably within the same trial. Early proof-of-concept monotherapy studies or single-center studies may be included as part of the phase II development program. Patients included in short-term studies should be allowed to enter long-term open-label follow-up for collection of additional safety information.

II.2. Short-term studies

II.2.i. Adjunctive therapy trial in patients with clinical indication for cardiac catheterization - IVUS

II.2.i.A. Objectives
To evaluate short-term efficacy and tolerability during adjunctive therapy use

II.2.i.B. Primary endpoints
a. The percent change in plaque volume (follow-up – baseline)/baseline x 100) measured by 3D IVUS.

II.2.i.C. Secondary endpoints
a. Absolute change in plaque volume on three dimensional (3-D) IVUS
b. Change in percent plaque volume on 3-D IVUS
c. Changes in plaque volume in anatomically comparable 5-mm segments centered on the sites with lowest and highest plaque burden at baseline by 3-D IVUS
d. Changes in plaque characterization indices assessed by IVUS
e. Coronary score assessed by quantitative coronary angiography (QCA) defined as the per-patient mean of the minimal lumen diameter for all lesions measured.
f. Incidence of adverse events

II.2.i.D. Exploratory endpoints
Changes from baseline in markers of biochemical target activation and markers of atherosclerosis pathophysiology (ie. inflammatory markers).

II.2.i.E. Study design
A multicenter, randomized, placebo-controlled, parallel-group design is typically used. Inclusion criteria usually require patients to be at least 18 years of age and with a need for diagnostic coronary angiography for clinical indication (with or without percutaneous coronary intervention). Patients are usually required to have stable background treatment following standard care practice. Patients are screened and the initial IVUS examination is performed in a target coronary artery which must not have undergone previous percutaneous coronary intervention (PCI) nor be a candidate for intervention at the time of the baseline catheterization. The baseline phase should be of sufficient duration to allow for scheduling of catheterization procedures within an acceptable time and stabilization of concomitant medication (4 weeks). Patients are then randomized to treatment groups including placebo and typically 2-3 doses of active study medication. A treatment phase of variable length (1-12 months) is initiated based on the characteristics of the drug and the mechanism of action. During the active treatment period, background medication must remain as stable as possible. The follow-up IVUS procedures are performed from 6 weeks to 12 months (see below for longer term IVUS studies) following the baseline IVUS.

II.2.i.F. Planned sample
Assuming that the standard deviation of the % change in plaque volume is around 6%, a sample size of approximately 65 patients per treatment group will provide 80% power to detect a 3% difference in plaque volume between the placebo group and one of the active treatment groups using a two-sided significance level of 0.05. Approximately 50 patients per treatment group are needed to detect a difference between the placebo group and two combined active groups in a 3-arm trial given the same assumptions. The final sample size must also account for loss-to-follow up of approximately 20 – 25%, given the cardiac catheterization procedures required.

II.2.i.G. Study population
Adults who require diagnostic coronary angiography for clinical indication, and stabilized to other background medications according to standard of care practices.

II.2.i.H. Specific inclusion criteria
a. Male or female aged \( \geq \) 18 years
b. Female patients who are not of childbearing potential (at least two years postmenopausal, surgically sterile, or practicing adequate contraception)
c. Scheduled for clinically indicated coronary angiography (with or without PCI)
d. Presence of at least one luminal diameter stenosis of 20% or more in one coronary artery by visual (angiographic) estimation
e. Presence of a non-PCI target coronary artery in which IVUS examination can be performed (target vessel)
f. The target vessel must not have undergone previous PCI nor be a candidate for intervention at the time of the baseline catheterization

II.2.i.I. Specific exclusion criteria
a. Previous or planned coronary artery bypass surgery
b. Any medical or psychiatric condition which in the opinion of the investigator makes the patient an unsuitable candidate for the study

c. History of alcohol or drug abuse within the past year

II.2.i.J. Tools for assessing primary endpoints

Assessment of the main outcome measures of atherosclerosis regression is performed by 3-D reconstruction of IVUS images. All the IVUS images should be interpreted by experienced technicians supervised by a cardiologist blinded to treatment assignment. The baseline and end-of-treatment studies should be viewed together. The use of reproducible IVUS landmarks (i.e. aorto-ostial junction, branches) and a known pullback speed (0.5 mm/sec) facilitate comparison of the same 30-mm segment on both studies and permit volumetric (3-D) analysis. Frame-by-frame review of the images is also systematically used to confirm matching of segments. The images are digitized and quantitative analysis performed. The lumen and external elastic membrane (EEM) borders can be traced manually or using an edge detection algorithm if all tracings are visually verified.

Plaque, lumen and total vessel volumes are computed by multiplying the corresponding areas of each of the cross-section by the distance between neighbouring slices and by then adding all the products. Cross-sections are analyzed in the 30-mm segment of interest at both baseline and follow-up. Plaque, lumen and total vessel volumes are first computed for the entire length (30 mm) of the analyzed segment. Volumes are also calculated on 5-mm segments centered on the sites with a) smallest plaque burden at baseline and b) largest plaque burden at baseline. In addition to the absolute plaque volume, percent plaque volume is also calculated as plaque volume divided by total vessel volume times 100. Detailed analysis of plaque composition is then also performed with IVUS at baseline and follow-up using plaque characterization indexes.

For the quantitative analysis of coronary angiograms (QCA), all angiograms from a given patient should be viewed together and analyzed by experienced technicians supervised by a cardiovascular radiologist blinded to the patients’ treatment assignments.

II.2.i.K. Specific criteria for early withdrawal and discontinuation

Standard criteria for early withdrawal include withdrawal of consent, or adverse events that render the continued treatment of patient(s) medically unacceptable.

II.2.i.L. Data analysis method

The data analysis methods used may be dependent on the objective of the trial. One method is to base the analysis of efficacy variables on the intention-to-treat (ITT) population (including all randomized patients who have follow-up data). However, for exploratory purposes, the per-protocol population (patients with no major protocol violations) may also be investigated. Statistical tests are generally two-sided and p values ≤0.05 are often considered statistically significant. Percent change is analyzed using analyses of variance (ANOVA) models. Analysis of covariance (ANCOVA) models adjusting for baseline values are used for absolute change. Coefficients of correlation are computed to assess the relationship between changes in inflammatory markers and primary and secondary endpoints. The safety analysis includes adverse events presented by treatment group using descriptive statistics (frequencies and counts).

II.2.ii. Adjunctive therapy trial in patients with stable atherosclerosis – brachial artery reactivity

II.2.ii.A. Objectives

To evaluate short-term efficacy and tolerability during adjunctive therapy use

II.2.ii.B. Primary endpoints

a. Change in flow mediated dilation (FMD) defined as the percent change from brachial artery diameter (mm) at rest to maximal brachial artery diameter during reactive hyperemia.
II.2.ii.C. Secondary endpoints
Secondary endpoints are designed to distinguish between vasodilatory changes resulting from changes in endothelial function as opposed to smooth muscle function.
   a. Change in nitroglycerin-mediated dilation (percent change from brachial artery diameter before administration to maximal brachial artery diameter after administration of sublingual nitroglycerin, an endothelium-independent vasodilator)
   b. Change from baseline in normalized brachial artery diameter
   c. Change from baseline in blood pressure, heart rate and velocity versus time integral

II.2.ii.D. Exploratory endpoints
   a. Changes from baseline in markers of biochemical target activation and markers of atherosclerosis pathophysiology (i.e. inflammatory markers).

II.2.ii.E. Study design
A randomized, placebo-controlled, parallel-group design is generally used. Single-center or small multi-center studies are often preferable due to inter-operator variability. There are typically three phases, a phase for stabilization of concomitant medication and diet, a single-blind placebo-baseline phase, and a double-blind treatment phase. Inclusion criteria usually require patients to be at least 18 years of age with objective evidence of coronary artery disease, peripheral artery disease or carotid artery disease. At screening, patients are asked to refrain from significant changes in dietary, smoking, and exercise habits for the duration of the trial as these may affect their endothelial function. Following the baseline period (typically 4 weeks), patients have vascular reactivity assessed through ultrasound procedures performed on the brachial artery. Patients are then randomized to treatment groups including placebo and typically 2-3 doses of active study medication. A treatment phase of 8-12 weeks is initiated based on the characteristics of the drug and the mechanism of action. During the active treatment period, background medication must remain as stable as possible. The follow-up brachial assessments are performed at the end of the treatment period.

II.2.ii.F. Planned sample
A sample size of approximately 45 patients per treatment group will provide 80% power to detect a 2.5% difference in change in FMD between the placebo group and one of the active treatment groups assuming that the standard deviation of the change is 4% and using a two-sided significance level of 0.05.

II.2.ii.G. Study population
Adults with objective evidence of coronary artery disease, peripheral artery disease or carotid artery disease and stabilized to other background medications according to standard of care practices.

II.2.ii.H. Specific inclusion criteria
   a. Males or females aged ≥18 years
   b. Female patients who are not of childbearing potential (at least two years postmenopausal)
   c. Other specific inclusion criteria are based on defining objective evidence of atherosclerotic disease as follows:
      • Stable angina pectoris for which frequency, severity, duration, time of appearance, and precipitating events have not changed for 60 days prior to screening;
      • Myocardial ischemia as evidenced by any of the following
         ✓ Stress ECG showing ischemic ST-segment response
         ✓ Stress echocardiography showing myocardial wall motion abnormality, or
         ✓ Myocardial perfusion scan showing a myocardial perfusion defect
      • At least 50% occlusion of the lumen of one or more coronary arteries, as evidenced by coronary angiography, IVUS, or other methods
      • ECG evidence of Q-wave myocardial infarction
• Iliac, femoral or carotid artery atherosclerosis as evidenced by angiography, ultrasound duplex scan, IVUS or other methods
• Previous carotid endarterectomy, peripheral bypass surgery, or abdominal aneurysm.

II.2.ii.I. Specific exclusion criteria
a. History of myocardial infarction or unstable angina within 4 weeks prior to screening
b. History of PCI, coronary artery bypass, cerebrovascular accident, or diagnosis of heart failure within 3 months prior to screening
c. Uncontrolled hypertension
d. Uncontrolled diabetes
e. Concomitant administration of supplemental antioxidants, L-arginine supplements, dipyridamole, pentoxifylline, angiotensin-converting enzyme (ACE) inhibitors, or ACE receptor inhibitors.
f. Any medical or psychiatric condition which in the opinion of the investigator makes the patient an unsuitable candidate for the study
g. History of alcohol or drug abuse within the past year

II.2.ii.J. Tools for assessing primary endpoints
High resolution B-mode ultrasound imaging is used to determine the diameter of the right brachial artery before and immediately after reactive dilation induced by ischemia which is produced by inflating a blood pressure cuff to 200 mmHg or to 50 mmHg greater than the systolic blood pressure (whichever is higher) for 5 minutes. Arterial diameter at baseline and at 60 seconds after cuff deflation is recorded. The relative difference is the measure of reactivity. Image analysis should be performed by a central laboratory with experienced technicians blinded to study treatment.

II.2.ii.K. Specific criteria for early withdrawal and discontinuation
The investigator will discontinue a patient before completion of the study if consent is withdrawn or if it is medically unacceptable to continue treatment due to adverse events.

II.2.ii.L. Data analysis method
The analysis of efficacy variables may be based on the intention-to-treat (ITT) population (including all randomized patients who have follow-up data). All statistical tests are two-sided and p values ≤0.05 are considered statistically significant. Analysis of variance or analysis of covariance adjusting for the baseline values could be used to compare the primary and secondary endpoints between groups. In any case, data analysis methods should be planned a priori and chosen to answer the specific objectives of the trial.

II.3. Long-term studies
Since the accumulation of lipid rich plaque material develops slowly over several decades, it has generally been accepted that chronic therapy would be required in order to detect the effects of these agents on measures of plaque burden or subsequent risk of cardiovascular events. For example, while some recent IVUS trials have demonstrated efficacy of exploratory agents when administered over a period of weeks, approved lipid-lowering agents with a demonstrated ability to decrease cardiovascular morbidity and mortality have generally required treatment durations of 18 months or more in order to demonstrate plaque volume changes detectable with IVUS or carotid ultrasound. Therefore, trials may need to be designed to evaluate long-term administration in order to detect significant treatment effects. The primary objective of these long-term studies is to provide an opportunity for the drug effect to be realized and to obtain data on tolerability and safety during long-term use. The studies described above as short-term studies may also be considered long-term studies depending on the study drug mechanism of action and the anticipated time required for treatment effect. In addition to the use of QCA and IVUS in Phase II long-term studies to detect anti-atherosclerotic effects, carotid imaging is also an effective imaging tool.
II.3.i. Adjunctive therapy trial in patients with clinical indication for cardiac catheterization - IVUS

II.3.i.A. Objective
To evaluate long-term efficacy and tolerability during adjunctive therapy use

Study description is similar to short-term study description, except that long-term studies using IVUS generally provide for follow-up evaluations after 12-24 months.

II.3.ii. Adjunctive therapy trial in patients with stable atherosclerosis – carotid intima-media thickness (IMT) evaluations

II.3.ii.A. Objective
To evaluate long-term efficacy and tolerability during adjunctive therapy use

II.3.ii.B. Primary endpoints
a. The change over time in mean maximum IMT across 12 pre-selected carotid arterial segments.

II.3.ii.C. Secondary endpoints
a. The difference in the slope of left and right common carotid artery far wall mean IMT progression (up to 200 far-wall measurements over a distance of approximately 2 cm beginning 1 cm below the bifurcation and using a standard angle of imaging)
b. The change in mean far wall IMT
c. The change in maximum far wall IMT
d. The rate of carotid artery progression measured as linear slope over annual ultrasound examinations
e. The average of the maximum carotid IMT of the far wall of up to 4 arterial segments (right and left distal common and right and left carotid bulb)
f. The rate of progression in the far wall of the left and right carotid bulb IMT
g. The rate of progression of the mean of maximum IMTs of the left and right common carotid artery
h. Change from baseline in mean IMT of the left and right common carotid arteries
i. Change from baseline in the maximum IMT of the left and right common carotid arteries
j. Incidence of adverse events.

II.3.ii.D. Exploratory endpoints
Changes from baseline in markers of biochemical target activation and markers of atherosclerosis pathophysiology (i.e. inflammatory markers)

II.3.ii.E. Study design
A multicenter, randomized, placebo-controlled, parallel-group design is generally used. Inclusion criteria usually require patients to be at least 18 years of age with evidence of carotid and coronary atherosclerosis. Patients are usually required to have stable background treatment following standard care recommendations. Patients are screened and the initial carotid ultrasound examination is performed bilaterally on the common and internal arteries. Patients are then randomized to treatment groups including placebo and typically 2-3 doses of active study medication. It is generally thought that treatment effects on the common carotid artery mean IMT could be seen after one year of treatment, but an 18-24 month treatment period would enhance the likelihood of this. Due to the non-invasive nature of carotid IMT assessment, serial examinations of the carotid wall are often undertaken.

II.3.ii.F. Planned sample
The sample size calculation is based on an average expected effect of 0.02 mm/yr change from baseline in carotid IMT measurements in at least one treatment group (for a total change of 0.04 mm in a 24-month
study). No change is expected in the placebo group. According to the literature, 0.20 mm would be a reasonable estimate of the standard deviation of the change in IMT from baseline to 24 months. Under these assumptions, a sample size of approximately 400 patients per treatment group would provide 80% power to detect a difference of 0.04 mm in change in IMT between the placebo group and one of the active treatment groups using a two-sided significance level of 0.05.

II.3.ii.G. Study population
Patients with evidence of atherosclerotic disease

II.3.ii.H. Specific inclusion criteria
a. Patients may be selected based on a pre-specified minimal baseline IMT measurement (> 0.8 mm),
b. Patients may be selected based on other evidence of atherosclerotic cardiovascular disease
c. Patients may be selected based on both of these criteria.

II.3.ii.I. Specific exclusion criteria
a. History of carotid revascularization
b. Patients in whom a screening IMT is suboptimal
c. Any medical or psychiatric condition which in the opinion of the investigator makes the patient an unsuitable candidate for the study
d. History of alcohol or drug abuse within the past year

II.3.ii.J. Tools for assessing primary endpoints
Images of the right and left common carotid and internal carotid arteries are captured, including images of the near and far wall, using high-resolution B-mode ultrasound. Ultrasound methodology should be specifically designed to include procedures to quality control the critical components of measurement variation including instrumentation, and ultrasound operations. Standardization of ultrasound machines at all sites is optimal, but not necessary. Image analysis is performed centrally at a center with experienced technicians.

II.3.ii.K. Data analysis method
The analysis of efficacy variables may be based on the intention-to-treat (ITT) population (including all randomized patients who have follow-up data). The continuous efficacy endpoints are analyzed using an analysis of covariance (ANCOVA) model with treatment and center as effect and the parameter’s baseline as covariate. If serial examinations of the carotid wall are undertaken, repeated measures ANOVA could also be conducted to study the way IMT changes over time. Chi-square tests or Cochran-Mantel-Haenszel tests using center as a stratification factor are used for categorical endpoints. Other data analysis methods more suitable for exploratory evaluations may be employed, depending on the objectives of the trial.

III. PHASE III STUDIES FOR REGISTRATION OF NEW ANTI-ATHEROSCLEROTIC DRUGS

III.1. Outline of a typical development plan
Phase III studies are conducted to establish the overall risk-benefit relationship of the drug and to provide adequate information for drug labeling. Therefore, for an atherosclerosis indication, the endpoints, patient population, and patient numbers should be consistent with these goals. It is important to recognize that no anti-atherosclerotic drugs working by a non-lipid mechanism have yet been approved. Additionally, there are no official guidelines written to direct the development of such agents. Three concerns have previously been identified for drugs working through novel, non-LDL cholesterol mechanisms. The first concern is that use of true placebo-controlled trials is not ethical in high-risk patients. Secondly, the administration of background “usual care” consisting of statins and other lipid-lowering therapy may add to the complexity of trial design and interpretation. Finally, with the need to demonstrate the effectiveness
of adjunctive therapy, sample size may become prohibitively large. The use of the imaging technologies outlined in Phase II trial descriptions may some day provide sufficient data to support the use of these endpoints as a surrogate for the reduction of cardiovascular risk. However, to date, the ability to rely solely on imaging endpoints in Phase III is yet theoretical. Therefore the development plan outlined in this document consists of a multicenter (usually multi-national), double-blind, randomized, placebo-controlled trial of two parallel groups designed to assess the combined incidence of cardiovascular morbidity and mortality.

III.2. Long-term adjunct therapy trial

III.2.i. Adjunctive therapy trial in patients at high risk for a major cardiovascular event

III.2.i.A. Objectives
To assess the effect of the investigational drug versus placebo on the combined incidence of cardiovascular morbidity and mortality

The definition of cardiovascular morbidity may include only “hard” endpoints such as non-fatal myocardial infarction, and stroke, or be expanded to include other “soft” endpoints, such as the need for coronary revascularization procedures, worsening angina requiring hospitalization, objective evidence of ischemia, and incidence of peripheral arterial disease.

III.2.i.B. Primary endpoint
a. Combined incidence of cardiovascular morbidity and mortality

III.2.i.C. Secondary endpoints
a. Incidence of all cause mortality
b. Incidence of cardiovascular mortality
c. Incidence of cardiovascular morbidity
d. Combined incidence of a subset of the events within the cardiovascular morbidity and mortality definitions.
e. Incidence of adverse events.

III.2.i.D. Exploratory endpoints
Changes from baseline in markers of biochemical target activation and markers of atherosclerosis pathophysiology (ie. inflammatory markers).

III.2.i.E. Study design
A multicenter, randomized, adjunctive therapy, double-blind, placebo-controlled parallel-group design could be utilized. Patients should receive standard care, including treatment of underlying diabetes, hypercholesterolemia and hypertension, prior to entry into this trial. There are several reasonable approaches to overall study design. One approach is to design the trial so that it will be complete with a fixed number of patients completing a minimal, fixed treatment period. Another approach is to design the trial so that it will complete when a predetermined number of patients have experienced a primary event based on expected control group rates and anticipated treatment effects.

III.2.i.F. Planned sample
The sample size will be based on the expected event rate in the selected patient population as well as the anticipated magnitude of treatment effect as demonstrated in Phase II trials. Other factors to consider in determining sample size include standard of care in country/region where trial is being conducted, planned trial duration, and time to effect based on an individual agent’s mechanism of action. A sample size of 2,000
– 6,000 moderate to high-risk patients per treatment group with minimally 18-month follow up would be a reasonable expectation.

III.2.i.G. Study population
Patients at moderate to high risk of cardiovascular morbidity and mortality based on the presence of one or more risk factors. Patients should receive treatment for modifiable risk factors prior to entry into trial.

III.2.i.H. Specific inclusion criteria
a. Specific inclusion criteria are selected based on identification of patients with moderate to high risk of cardiovascular morbidity and/or mortality.

b. Inclusion criteria may include one or more of the following major risk factors:
   - Cigarette smoking
   - Hypertension (BP > or = 140/90 mm Hg)
   - Low HDL cholesterol (< 40 mg/dL)
   - Family history of premature CHD (CHD in male first degree relative < 55 years; CHD in female first degree relative < 65 years)
   - Age (≥ 55 years)
   - Diabetes

b. Inclusion criteria may include one or more of the following life-habit risk factors
   - Obesity
   - Physical inactivity
   - Atherogenic diet

c. Inclusion criteria may include one or more of the following emerging risk factors
   - C reactive protein
   - Lipoprotein (a)
   - Homocysteine
   - Prothrombotic and proinflammatory factors
   - Impaired fasting glucose
   - Evidence of subclinical atherosclerotic disease

d. Inclusion criteria may include one or more of the following life-habit risk factors
   - C reactive protein
   - Lipoprotein (a)
   - Homocysteine
   - Prothrombotic and proinflammatory factors
   - Impaired fasting glucose
   - Evidence of subclinical atherosclerotic disease

t. Specific exclusion criteria
a. Any medical or psychiatric condition which in the opinion of the investigator makes the patient an unsuitable candidate for the study

b. History of alcohol or drug abuse within the past year

III.2.i.J. Tools for assessing primary endpoints
An independent Clinical Endpoint Committee (CEC) is established to review and classify all suspected major cardiovascular events according to pre-defined guidelines and definitions. The CEC remains blinded to treatment group assignment for the duration of the trial. The CEC is typically composed of 5 cardiologists experienced in patient care. Reviewers are assigned at random to review each case and classify the event. For the classification to be considered final, consensus between two primary reviewers is required. If consensus between two primary reviewers is not reached, the case is typically decided by simple majority of the entire committee.

III.2.i.K. Specific criteria for early withdrawal and discontinuation
The investigator will discontinue a patient before completion of the study if consent is withdrawn or if it is medically unacceptable to continue treatment due to adverse events. An independent Data Safety Monitoring Board (DSMB) is typically established and responsible for assessing patient safety during the course of the trial. A trial may be discontinued if such a recommendation is made by the DSMB based on a periodic review of the generated data.
III.2.i.L. Data analysis method

The analysis of efficacy variables is based on the intention-to-treat (ITT) population. Statistical tests are 2-sided at the 0.05 level of significance. Survival analysis including Kaplan-Meier curves and log-rank tests to compare survival curves across groups are used. Multivariate analysis using Cox proportional hazards models may also be performed.

IV. OTHER STUDIES (SPECIAL INDICATIONS)

Other studies may be considered for inclusion in the development plan for anti-atherosclerotic agents that work by non-lipid mechanisms. Patient populations that may warrant focused evaluation in separate clinical trials include patients with traditional or emerging risk factors, or genetically defined conditions which confer high risk of CAD. Separate trials in patients with diabetes, metabolic syndrome, genetic predisposition to CAD, peripheral arterial disease, etc. may provide focused and useful information for further development. There may be also other imaging technologies that may have an important role in drug development aside from those mentioned previously. Magnetic resonance imaging, computed tomography, positron emission tomography and other technologies may all have an important role in determining the effectiveness of potential anti-atherosclerotic agents. These additional studies may have utility in early stage development to help establish proof of concept and better define the mechanism of action. Alternately further evaluations of specific populations and technologies may have greatest utility in Phase IV trials to further define established efficacy.

V. EXAMPLES OF LANDMARK WELL DESIGNED TRIALS

Intravascular Ultrasound

Brachial artery reactivity

Carotid intima-media thickness (IMT)

Major cardiovascular morbidity and mortality studies

VI. SUGGESTED READINGS

Chapter 14. Drugs for Heart Failure

Anique Ducharme, M.D., M.Sc. 1

Simon de Denus, B. Pharm, M.Sc. 2

1 Montreal Heart Institute
5000 Belanger Street East,
Montreal, Quebec  HIT 1C8
CANADA
Tel 514-376-3330 ext.3927
Fax 514-593-2575
Email ducharme@icm.umontreal.ca

2 Faculty of Pharmacy
University of Montreal
Montreal, Québec
CANADA
and
Department of Pharmacy
Montreal Heart Institute
Montreal, Québec
CANADA

Corresponding Author:
Anique Ducharme, M.D., M.Sc.
I. INTRODUCTORY REMARKS

Heart failure (HF) has reached epidemic proportions in the U.S. with nearly 5 millions individuals afflicted with this condition. Despite major advances in the treatment of this condition, both pharmacological and non-pharmacological, HF is responsible for more than 250 000 deaths per year in the US alone. Furthermore, this condition is associated with significant morbidity with almost 1 million hospital admissions each year. Because of this, the management of HF is associated with an annual cost of more than 24 billion dollars. These statistics clearly illustrates that novel therapies, both pharmacological and non-pharmacological, are still required to improve the prognosis of these patients. Unfortunately, the development of new HF drugs is complicated by a lack of established surrogate markers as reliable as blood pressure in hypertension or cholesterol concentrations in dyslipidemia.

II. PHASE II STUDY TO EVALUATE THE EFFECTIVENESS AND SAFETY OF NEW HEART FAILURE DRUGS

II.1. Outline of a typical development plan

This phase generally begins with open, single-blind or double-blind studies comparing various dosages of a new agent in addition to or compared to an established treatment. Frequently, these dosages are derived from the hypertension literature, because these dosages have been proven to be safe and effective. These studies are generally conducted in a limited number of centers.

Results of these pilot studies are vital for appropriate design of key studies, since these small initial trials then lead to large dose-finding multicenter, controlled trials. These studies are generally conducted using a parallel design and compare the investigational drug with placebo or an active-control, depending if it is expected to be used in addition to current treatment or as an alternative to an established treatment. For example, angiotensin receptor blockers (ARBs) were initially compared to angiotensin-converting enzyme (ACE) inhibitors because they were expected to provide a superior clinical benefit through a more complete inhibition of the effects of the renin-angiotensin-aldosterone system, while being better tolerated because they do not cause an accumulation of bradykinin. Oppositely, when experimental data demonstrated that bradykinin accumulation could positively influence the evolution of HF, the addition of an ARB in patients already receiving an ACE inhibitor was investigated with positive results. These dose-finding studies will generally involve patients with HF of moderate to severe symptoms of New York Heart Association (NYHA) II-IV. Asymptomatic patients are usually excluded (NYHA functional class I). End points commonly consist of vital signs, exercise tolerance, hemodynamic measurements, echocardiographic measurements, neurohormone concentrations and patient’s assessment of well-being.

The potential for drug interactions with commonly used cardiovascular drugs with a high potential for interaction (warfarin, digoxin) and other drugs known to affect the metabolism of commonly used drugs (cimetidine, rifampin) may also be investigated, but these investigations are generally conducted in healthy or hypertensive patients and, unfortunately, only rarely in HF patients. Because the pharmacokinetics of some agents vary between healthy individuals and patients with HF, it is possible that the importance of a drug interaction could vary between these two patient populations and it would be therefore preferable to evaluate drug interactions of interest in the HF patient population also.

II.2. Short-term studies

II.2.A. Objectives

Evaluate both the short-term efficacy, generally through hemodynamic monitoring, neurohormonal measurements and functional class, and the safety of the agent.
II.2.B. Primary endpoints
Because there is no established surrogate marker to evaluate the efficacy of drugs in HF, endpoints vary greatly in short term studies, but generally consist of either:

- a. Hemodynamic measurements (filling pressures, cardiac index, systemic vascular resistance, blood pressure, heart rate)
- b. Improvement in symptoms or functional class
- c. Improvement in exercise capacity (peak oxygen consumption, sub maximal exercise tolerance or 6 minutes walk test).
- d. Measurements of neurohormones: catecholamines, brain natriuretic peptide (BNP), others.
- e. Echocardiographic or Magnetic Resonance Imaging (MRI) parameters: left ventricular (LV) dimensions, volumes, LV ejection fraction (LVEF), and degree of mitral regurgitation.
- f. Patients’ perception of dyspnoea and well-being.

II.2.C. Secondary endpoints
In these initial phase II studies, secondary endpoints will consist of the incidence of adverse drug effects (safety and tolerability) and mortality, cardiovascular events (CV mortality, myocardial infarction, stroke and cause-specific hospitalizations). The impact of the agent under investigation on quality of life may also be evaluated.

II.2.D. Exploratory endpoints
Relationship between dose, efficacy, and adverse effects. Drug withdrawal and rebound effects.

II.2.E. Study design
A least one randomized, double-blinded, parallel-group design, using multiple centers is generally conducted. Patients can be randomized to different doses of the investigated drug versus placebo or an active control, usually with a dose-ranging design. Patients should be receiving an optimal medical therapy at proven or maximally tolerated doses at the beginning of the study; they should be maintained on this regimen throughout the duration of the study. The drug titration period will depend on the agent being investigated and the need to monitor for vital signs, electrolytes, renal and liver function tests. Maintenance phase should be of at least 3 months duration, but no longer than 6 months. At the end of the study, the patients should be withdrawn from the investigational drug based on a pre-established schedule. If the drug is planned to be on the market in a near future, an open label extension phase can be offered to the patient.

Concomitant therapy
Unless contraindicated, all patients should be on optimal medical therapy (both agents and doses) consisting of agents from the following class of agents:

- a. ACE inhibitors or ARBs
- b. Beta-blockers
- c. Spironolactone (in NYHA class III-IV patients)
- d. Diuretics, usually
- e. Digoxin

II.2.F. Planned sample
In these initial trials, the sample size will depend on the end points measured. Typically the smaller initial trials will include approximately 50 to 100 patients, whereas the larger dose-finding trial can include up to a few hundred patients.

II.2.G. Study population
Patients with symptomatic, chronic HF
II.2.H. Specific inclusion criteria
   a. Adults (at least 18 years of age)
   b. Left ventricular ejection fraction ≤40% (or 35%), measured in the last 6 months.
   c. NYHA class II-IV symptoms.
   d. Stable symptoms: duration of stability varies according to study (4 days-3 months, except studies of decompensated HF)
   e. Stable medical regimen ≥1 month (except diuretics)

II.2.I. Specific exclusion criteria
   a. Inability to provide informed consent
   b. Unstable angina, myocardial infarction or coronary revascularisation within the last 6 weeks-3 months.
   c. Planned cardiac surgery (within 3-6 months).
   d. Significant valvular disease.
   e. Systolic blood pressure < 90 mmHg (except in studies of cardiogenic shock).
   f. Uncontrolled hypertension (blood pressure systolic >170 mm Hg or diastolic >105 mm Hg).
   g. Heart rate < 60 bpm for agent causing significant bradycardia.
   h. Advanced heart block without pacemaker.
   i. Significant renal insufficiency (definition varies between trials and the nature of the drug studied).
   j. Significant non-cardiac co-morbidities or laboratory abnormalities.
   k. Pregnancy and women of childbearing potential unless a safe contraception method is used.
   l. A potentially reversible cause of HF (e.g. thyrotoxicosis or uncontrolled supraventricular arrhythmia).
   m. Known drug or alcohol misuse, poor compliance with treatment or any other serious systemic disease that might complicate management and reduce life expectancy.
   n. Administration of any investigational drug within the preceding 30 days.

II.2.J. Tools for assessing endpoints
Primary endpoints
   a. Echocardiography or MRI
   b. Exercise stress test with or without O2 consumption measurements
   c. Right heart catheter
   d. Blood samples
   e. Visual analog scale (for symptoms’ self-evaluation)

Secondary endpoints
   a. Emergency room visits log
   b. Discharge summary: cause & number of hospital admission
   c. Death certificate
   d. Quality of life questionnaire

The measure of the endpoints will be performed according:
   a. Review of the Case Report forms for all clinical events, including queries of the patient’s clinical chart if needed
   b. Standardized method of evaluation: Core laboratory for consistency of data: echocardiography, MRI, blood samples (neuro-hormones, BNP, etc.), and exercise stress test.

II.2.K. Specific criteria for early withdrawal and discontinuation
   a. Patient’s request/consent withdrawal
   b. Pregnancy
c. Serious adverse effects
d. When, in the opinion of the physician, continuation of the therapy is not in the best interest of the patient.

II.2.L. Data analysis method
The analysis of efficacy variables is based on the intention-to-treat principle, which includes all randomized patients. The specific statistical analyses will vary according to the types of variables measured and end points evaluated. All statistical tests are two-sided and statistical significance is generally established with a $p$ value $\leq 0.05$. Multiple statistical methods can be used for the primary and secondary end points.

II.3. Long-term studies
At the end of the randomized phase, it is usual for patients completing short-term studies to be offered an extension phase with open-label follow-up. The objective of this prolongation is to provide data on tolerability and safety during long-term use. The study drug may be continued for as long as it is felt to be clinically beneficial, until the agent is available on the market.

III. PHASE III STUDY TO EVALUATE THE EFFECTIVENESS AND SAFETY OF NEW DRUGS FOR HEART FAILURE

III.1. Outline of a typical development plan
Whereas phase II studies generally focus on surrogate markers of efficacy, phase III trials are specifically designed to evaluate the impact of the drug under investigation on clinical outcomes, mainly death and hospitalization. This phase generally includes at least one large multicenter double-blind, randomized placebo or active control trial of patients with symptomatic HF (NYHA functional class II-IV). Patients with NYHA functional class I are generally excluded because of their lower risk of cardiovascular events. For patients in NYHA II, a criteria for potential instability is often required (ex. hospital admission for HF within 6-12 months or use of diuretics). The follow-up of these studies should be of at least 6 months, preferably 12 months, but can be extended to several years.

III.2. Short-term studies
The design for phase III trials in HF is similar to that described above for phase II trials. Sometimes, multiple doses or regimens are explored.

III.3. Long-term studies

III.3.A. Objectives
To evaluate the long-term efficacy and safety of a given treatment compared with optimal medical management.

III.3.B. Primary endpoints
Mortality (total, cardiovascular), or a combination of mortality and morbidity (which generally consists of cardiovascular hospitalizations or episodes of worsening HF) is usually used.

III.3.C. Secondary endpoints
Secondary endpoints for phase III trials will consist of any of the previously mentioned endpoints not included in the primary endpoints or any components of a combined end point. In addition, drug safety and tolerance is reported as a secondary endpoint. Furthermore, the following can also constitute secondary endpoints: ischemic events (myocardial infarction, angina, need for revascularisation), changes in neurohormones or inflammatory markers levels, exercise tolerance, quality of life or changes in
echocardiographic (or MRI) parameters (left ventricular dimensions, LV ejection fraction [LVEF], ventricular volumes, degree of mitral regurgitation, etc…).

III.3.D. Exploratory endpoints
Exploratory analyses are generally limited to any sub-group analysis not prospectively defined.

III.3.E. Study design
Phase III study are usually multicenter, randomized, double-blinded placebo-controlled or active-controlled studies.

Concomitant therapy
Optimal medical treatment as described for the phase II in section II.2.E.

III.3.F. Planned sample
In phase III trials, the sample size is calculated based on the number of events expected in the population studied during the course of the trial and the expected reduction on these events rate with the new treatment. Usually, several thousand patients are planned to be enrolled.

III.3.G. Study population
As described for phase II studies in the section II.2.G.

III.3.H. Specific inclusion criteria
The inclusion criteria are generally the same as those used for the phase II.

III.3.I. Specific exclusion criteria
The exclusion criteria are generally the same as those used for the phase II. In addition, patients awaiting a cardiac transplantation are often excluded.

III.3.J. Tools for assessing endpoints
Primary endpoints
a. Clinical end-points committee (CEC): an independent CEC is usually formed to review every event occurring during the course of the study, with blind adjudication.

Secondary end points
a. Discharge summary (cause & duration of hospitalization)
   b. Worsening HF: chest X-Ray, BNP level, IV diuretic used or number of unplanned HF clinic visits.

The measurement of endpoints is performed as described for the phase II studies in the section II.2.J.

III.3.K. Specific criteria for early withdrawal and discontinuation
These criteria are generally the same as those discussed for phase II trials in section II.2.K.

III.3.L. Data analysis method
The data analysis is performed in an intention to treat fashion although an “on treatment” analysis is sometimes performed retrospectively in some cases (e.g. high rate of withdrawal).

IV. PHASE III STUDY TO EVALUATE THE EFFECTIVENESS AND SAFETY OF NEW DRUGS FOR HEART FAILURE IN SPECIAL POPULATIONS
The majority of HF trials have previously focused on patients with depressed left ventricular systolic function (ex. LVEF < 40%), probably because they have a poorer prognosis than patients with symptomatic HF and
preserved systolic function. Nevertheless, this preserved HF group represents almost a third of all HF patients, but has not been studied extensively. With the aging of the population and the prevalence of diabetes, which has reached epidemic proportions, this group will increase dramatically. Thus, efforts have been put forward lately to study more carefully HF patients with preserved systolic function. Unfortunately, the diagnosis is often difficult, since dyspnoea is often multifactorial (including participation of cardiac causes, hypertension and pulmonary disease). Furthermore, there is a lack of consensus on the criteria to define and investigate this entity. Fortunately, a potential surrogate marker could help differentiate between shortness of breath of pulmonary or cardiac origin, the BNP, which is a hormone secreted mainly by the ventricle in response to stretch.

**IV.1. Outline of a typical development plan**

Similar to phase III study as outlined in section III.1, except for the inclusion criteria.

**IV.2. Long term studies**

**IV.2.A. Objectives**

As described for phase III studies in section III.3.A.

**IV.2.B. Primary endpoints**

As described for phase III studies in section III.3.B.

**IV.2.C. Secondary endpoints**

As described for phase III studies in section III.3.C.

**IV.2.D. Study design**

As described for phase III studies in section III.3.E.

**Concomitant therapy**

Optimal medical treatment is unknown in this patients’ group, and usually relies on the treatment of concomitant illnesses (ex. ischemia, dyslipidemia, diabetes, hypertension, etc.).

**IV.2.E. Planned sample**

As for phase III trials of HF patients with systolic dysfunction, the sample size is calculated based on the number of events expected in the population studied during the course of the study and the expected reduction on these events rate with the new treatment. Unfortunately, since HF with preserved systolic function has not been fully studied, assumptions on prognosis and number of events are currently difficult to assess.

**IV.2.F. Study population**

a. Preserved LV systolic function (EF > 40%)

b. Others, as outline for phase II studies in section II.2.G.

**IV.2.G. Specific inclusion criteria**

The inclusion criteria are similar to those described for phase II studies in section II.2.H.

**IV.2.H. Specific exclusion criteria**

The exclusion criteria are similar to those described for phase II studies in section II.2.I.

**IV.2.I. Tools for assessing endpoints**

The tools to assess and measure the endpoints are usually the same as those used during the phase III in section III.3.J.
IV.2.J. Specific criteria for early withdrawal and discontinuation
Usually the same as those discussed for phase II trials in section II.2.K.

IV.2.K. Data analysis method
Usually the same as those used for the phase III in section III.3.L., but special emphasis should be put on the number of events needed (see above).

V. PHASE IV STUDIES
Phase IV studies are usually large-scale, open-label postmarketing surveillance trials. These studies take place once the FDA or EMEA has approved the agent. The goal is to evaluate if the safety and efficacy of the agent evaluated in controlled studies is maintained in a large population. These usually include patients from subgroups potentially at higher risk for adverse events. Physicians systematically document their observations concerning patients with HF on case report forms. This trial design is rarely used in the HF population.

V.1. Outline of a typical development plan
Similar to phase II and III studies outlined above.

V.2. Short-term study

V.2.A. Objectives
The goal is to find if the safety and efficacy of the pharmacological agent as evaluated in controlled studies is maintained in a large population. The duration is usually 6 months.

V.2.B. Primary endpoints
Similar to phase II and III studies outlined in section II.2.B. and III.3.B. In addition, global tolerability is rated among specific subgroups of patients: genders, age and co-morbidities. Quality of life questionnaire can also be included.

V.2.C. Secondary endpoints
Similar to phase II and III studies outlined in sections II.2.C. and III.3.C.

V.2.D. Study design
Randomized, active-controlled, open-label study

Concomitant therapy
As described in the phase II outline in section II.2.E.

V.2.E. Planned sample
While for antihypertensive drugs several thousand patients are usually necessary, the study of drugs indicated for HF usually include smaller cohorts. The number of patients enrolled in these trials have been as small as 50 and, generally, is limited to no more than a few hundred patients.

V.2.F. Study population
As described in the phase II studies outlined in section II.2.G.

V.2.G. Specific inclusion criteria
As described in the phase II studies outlined in section II.2.H.
V.2.H. Specific exclusion criteria
As described in the phase II studies outlined in section II.2.I.

V.2.I. Tools for assessing endpoints
   a. Tolerability questionnaire
   b. Blood samples
   c. Exercise tolerance test

Measure of endpoints
As described in the phase II outlined in section II.2.J.

V.2.J. Specific criteria for early withdrawal and discontinuation
As described in the phase II outlined in section II.2.K.

V.2.K. Data analysis method
As described in the phase II outlined in section II.2.L.

V.3. Long-term studies

V.3.A. Objectives
The goal is to compare the effects of drug A and drug B on clinical outcome. The duration may be up to 5 years. The only example of a phase IV study in patients with HF is the recently published COMET (Carvedilol Or Metoprolol European Trial). Since beta-blockers have been shown to reduce mortality in HF patients with systolic dysfunction, they aimed to compare the effects of carvedilol and metoprolol on clinical outcome, in patients who were on background treatment with diuretics and ACE inhibitors. The methods of the COMET trial will be used as an example of phase IV study.

V.3.B. Primary endpoints
The primary end points of the COMET study were all-cause mortality and the composite end point of all-cause mortality or all-cause admission.

V.3.C. Secondary endpoints
Similar to the ones described for phase III trial.

V.3.D. Study design
Randomized, active-controlled, open-label or double-blind study (COMET).

Concomitant therapy
   a. On stable HF treatment with ACE inhibitors for ≥4 weeks unless contraindicated.
   b. On treatment with diuretics (≥40 mg of furosemide or equivalent) for at least 2 weeks.
   c. Digitalis, ARBs, or other vasodilators could be used at the discretion of the investigators.

V.3.E. Planned sample
The COMET study was planned as an event-driven parallel-group survival study to compare carvedilol and metoprolol with respect to all-cause mortality. A total of 1020 fatal events were needed to detect a risk reduction of 20% with at least 80% power with an overall type I error of 0.05.

V.3.F. Study population
Eligible patients were men or women with symptomatic chronic HF (New York Heart Association [NYHA] class II–IV),
V.3.G. Specific inclusion criteria
   a. At least one cardiovascular admission during the previous 2 years.
   b. LV ejection fraction \( \leq 0.35 \), measured within the previous 3 months or left-ventricular end diastolic diameter \( \geq 6.0 \) cm and a fractional shortening < 20% measured by echocardiography.

V.3.H. Specific exclusion criteria
   a. A recent change of treatment, current treatment with diltiazem or verapamil, amiodarone (>200 mg per day) or class-I antiarrhythmic drugs.
   b. Patients with unstable angina, myocardial infarction, or coronary revascularisation or stroke within the previous 2 months.
   c. Uncontrolled hypertension (blood pressure systolic >170 mm Hg or diastolic >105 mm Hg).
   d. Hemodynamically significant valvular disease.
   e. Symptomatic and sustained ventricular arrhythmias within the past 2 months not adequately treated with antiarrhythmic drugs or defibrillator.
   f. Pregnancy, women with childbearing potential on inadequate contraception.
   g. Known drug or alcohol misuse, poor compliance with treatment.
   h. Any other serious systemic disease that might complicate management and reduce life expectancy.
   i. Patients in whom there was a contraindication to use of \( \beta \)-blockers, such as resting heart rate < 60 bpm, sick sinus syndrome, bifascicular block, second or third degree atrioventricular block unless treated with a pacemaker, sitting systolic blood pressure < 85 mm Hg, history of asthma or chronic obstructive pulmonary disease, peripheral arterial disease with symptoms at rest, or unstable insulin-dependent diabetes mellitus.

V.3.I. Tools for assessing primary endpoints
As described in the phase III outlined in section III.3.J.

Measure of end points
As described in the phase III outlined in section III.3.J.

V.3.J. Specific criteria for early withdrawal and discontinuation
As described in the phase II outlined in section II.2.K.

V.3.K. Data analysis method
Analysis was done by intention to treat.

V.4. Comments

Although the phase IV study design is common in clinical research focusing on hypertension, it is infrequently conducted in the HF population. Only one large phase IV study including patients with HF has been published, which is described in detail above.

VI. EXAMPLES OF LANDMARK WELL DESIGNED TRIALS

Phase II studies
Phase III studies

Phase III – Special population (HF with preserved LVEF)

Phase IV

VII. SUGGESTED READINGS
Chapter 15. Arrhythmias

Teresa Kus, M.D., Ph.D.

Associate Professor
Department of Pharmacology
Université de Montréal
Centre de Recherche
Hôpital du Sacré Cœur
5400, boulevard Gouin Ouest
Montréal, Québec H4J-1C5
CANADA
I. INTRODUCTORY REMARKS

Historical Perspective
The development of new antiarrhythmic drugs has become a major challenge since the publication of the Cardiac Arrhythmia Suppression Trial (CAST) results in 1989 (1). This was the first study designed to demonstrate that antiarrhythmic drugs with a potent Class I effect (flecainide and encainide) decrease mortality as compared to placebo in patients with remote myocardial infarction (MI) and frequent ventricular ectopic beats and who are therefore at increased risk for sudden death (SD) due to malignant ventricular arrhythmia. Improved survival was expected as these drugs markedly suppressed premature ventricular contractions in the same study patients. Instead, an increase in mortality was observed on antiarrhythmic drug compared to placebo (2,3).

This result affected antiarrhythmic drug development in two ways. First, the search for more potent drugs with Class I effect, i.e., slowing of conduction velocity by blockade of the sodium channel, was abandoned. In the setting of prior MI or ongoing ischemia, Class I effect was now generally accepted as potentially proarrhythmic. Instead, attention was directed to drugs which prolong action potential duration and refractoriness. In the presence of such a Class III effect, an arrhythmia dependent on a rapidly circulating electrical impulse would be interrupted when the wavefront encountered refractory tissue. The second major result of CAST was the perceived necessity of demonstrating the safety of such drugs in large placebo-controlled trials in patients with underlying cardiac pathology and left ventricular (LV) systolic dysfunction. This also proved to be important for the Class III drugs (3). In susceptible populations or under particular clinical circumstances, these drugs can induce excessive and heterogeneous prolongation of action potential duration with the consequent appearance of triggered automaticity. Such abnormalities can then set the conditions for torsade des pointes, a potentially lethal ventricular arrhythmia (4).

Present Indications for Antiarrhythmic Drug Therapy
Currently, the major indications for antiarrhythmic drug use are atrial fibrillation (AF) and atrial flutter (AFl). AF is a rapid (>400 bpm), disorganized, multi-circuit atrial rhythm whose prevalence increases with age (up to 8% in those over 80 years). It is responsible for approximately one-third of arrhythmia admissions and increases mortality two-fold and the risk of stroke five-fold (5). Acutely, it can manifest with symptoms of palpitations, hypotension, angina or heart failure. Further, if the ventricular response rate remains chronically elevated, it can result in a tachycardia-induced cardiomyopathy (CM). AFl, a single organized circuit usually in the right atrium, is less common but also increases the risk of cardiogenic emboli and provokes similar symptoms. Antiarrhythmic drugs are used acutely as an alternative to electrical cardioversion for both persistent AF and AFl so as to achieve sinus rhythm (SR) and chronically to prevent recurrence. They are also used to diminish the frequency of paroxysmal or spontaneously terminating episodes of AF and AFl. The currently available antiarrhythmics, however, are only moderately effective (30-60%), only a few are safe to administer in the presence of heart failure (dofetilide, amiodarone), and, in the case of the most effective (amiodarone), risk significant organ toxicity (6).

The use of antiarrhythmic drugs for long-term prevention of other supraventricular tachycardias (SVT) such as AV nodal reentry, reentry using an accessory pathway, and, in some centers where the expertise is available, atrial tachycardia and AFl, has taken a distant second place to catheter ablation of the critical component of the arrhythmia circuit. Antiarrhythmic drugs are now mainly used for these arrhythmias only while awaiting definitive curative ablation (7).

Pharmacological prevention of ventricular arrhythmias has also been dramatically challenged. Implantable defibrillators have been demonstrated to be superior to antiarrhythmic drugs in diminishing mortality in survivors of SD secondary to ventricular tachycardia (VT) or fibrillation in the presence of ischemic and non-ischemic CM and impaired systolic LV function (ejection fraction (EF) < 35%) (8,9). In this setting, however, antiarrhythmics can still play an adjunctive role by diminishing the number of episodes of ventricular arrhythmia or by slowing the rate of VT so as to allow painless pacing termination by the
defibrillator. Either effect should ultimately reduce the number of shocks delivered, improve the quality of life (QOL) of the patient and maximize the battery life of the defibrillator. When VT is well tolerated and LV function is relatively well-preserved (EF > 40%), long-term antiarrhythmic therapy with a drug with class III effect (sotalol or amiodarone) instead of defibrillator implantation can still be envisaged (8,9).

Antiarrhythmic drugs have so far not been successful when used as primary prevention of ventricular arrhythmias in patients at risk of SD. In general, drugs with Class I effect should not be given in either ischemic or non-ischemic CM because of increased mortality on drug (10). At best, the drugs studied to date with Class III effects have neutral effects on all-cause mortality in the post-MI patient population with congestive heart failure (CHF) (11). This latter result is in fact hypothesized to be due to a balance between an antiarrhythmic effect and a ventricular proarrhythmic effect (sotalol, dofetilide, azimilide) or the occurrence of serious organ toxicity (amiodarone). In this setting, too, the prophylactic implantation of defibrillators has proven more effective in reducing mortality compared to both placebo and amiodarone when LV systolic function is significantly reduced (EF < 35%) (12). Unfortunately, it is in the setting of CHF that the incidence of AF is highest and contributes significantly to a worse outcome. Therefore, it is necessary to demonstrate at least a neutral effect of drug on mortality in this population in order to allow its administration for atrial arrhythmias in such patients.

Clearly, there is a need for drugs which are more effective and safer in terms of proarrhythmic potential and organ toxicity than those currently available. Phase II studies should therefore compare the efficacy and safety of investigational drug X not only versus placebo but also against reference drug Y currently used in clinical practice. A pharmacokinetic profile allowing once daily dosing as well as minimal interaction with other drugs administered in such clinical settings would also be highly desirable.

II. PHASE II STUDIES TO EVALUATE THE EFFICACY AND SAFETY OF NEW ANTIARRHYTHMIC DRUGS IN ATRIAL FIBRILLATION AND ATRIAL FLUTTER

II.1. Outline of a typical development plan
Phase II studies of antiarrhythmic drugs are primarily directed at demonstrating efficacy and safety in the treatment of the most common atrial arrhythmias, AF and AFl. Various tools are available to document arrhythmia including the 12-lead electrocardiogram (ECG), Holter (24-hour) recording of 2 or 3 leads, and transtelephonic monitoring (TTM) or loop recording. The TTM loop recorder is especially useful in documenting transient episodes of arrhythmia. This can be done either automatically using preestablished criteria (rate and irregularity) for AF or by patient activation during a symptomatic episode. The device then stores in memory the preceding preprogrammed duration of recorded beats as well as an independently preprogrammed duration after activation of the storage function. The stored data can then be transmitted by telephone to a receiving station where symptoms are documented and the rhythm is interpreted.

In addition to arrhythmia documentation, the analysis of drug effect in Phase II studies also involves a detailed observation of effects on various ECG measures reflecting drug action on the sinus node (PP interval), ventricular response rate in AF or AFl, AV node conduction (PR interval) in SR, intraventricular conduction (QRS interval) and ventricular repolarization time in absolute terms (QT interval) as well as when corrected for heart rate (QTc interval). This latter measure is particularly important as an indicator of appropriate or excessive Class III effect. Other laboratory investigations are directed at modifiers of drug effect (electrolyte levels) or drug pharmacokinetics (eg., renal function tests) or at indicators of potential drug toxicity (eg., liver function tests, hematological values).

II.2. Short-term study
This trial is a multicenter, randomized, double-blind, placebo-controlled, parallel group study of the efficacy and safety of intravenous bolus administration of antiarrhythmic drug X versus reference drug Y in the acute termination of atrial fibrillation and flutter.
This study is often the first type of study done as it involves only single acute intravenous (IV) administration of drug and therefore requires only short-term monitoring of the subject. It incorporates dose ranging to search for effective plasma concentrations and doses and assesses the safety of these doses. Variations of such a study (with or without a reference drug) have been performed in the development of such Class III drugs as dofetilide (13) and ibutilide (14). It should be emphasized that the usual initial clinical approach to a patient presenting with AF/AFl should be followed, i.e., anticoagulation as recommended by current clinical guidelines and control of the ventricular response rate (6) before initiating study drug.

II.2.A. Objectives
Primary Objectives:
   a. To determine the efficacy of bolus I.V. administration of investigational drug X at the highest safe dose compared to placebo and to reference drug Y in achieving conversion of AF or AFl to SR (expressed as percentage conversion within T hours of I.V. dose).
   b. To determine the safety of investigational drug X at the most effective dose compared to placebo in subjects with AF or AFl.

Secondary Objectives:
   a. To determine the efficacy and safety of I.V. investigational drug X at doses A, B, C and D compared to placebo in achieving conversion of AF or AFl to SR.
   b. To assess the effects of doses A, B, C or D of drug X and of reference drug Y on ECG measures (sinus rate, and PR, QRS, QT and QTc intervals in responders and QRS, QT and QTc intervals and ventricular response rate in AF or AFl in non-responders).
   c. To determine the pharmacokinetics of investigational drug X administered I.V. in patients with AF or AFl and its modification by concomitant cardiac or minor to moderate hepatic and renal impairment.

II.2.B Primary endpoints
   a. Incidence of conversion to SR within time T of study drug administration.
   b. Incidence of side effects.

II.2.C. Secondary endpoints
   a. Mean time to conversion.
   b. Ventricular rate in non-responders on drug treatment with respect to baseline.

II.2.D. Study Design
Subjects who meet the study inclusion/exclusion criteria will be randomized in double-blind fashion to investigational drug X at dose A, B, C or D, the reference drug Y or placebo with all to be administered as a bolus I.V. In order to assure safety of the study drug at the doses proposed, such a study could be performed in a tiered fashion:
Tier 1: Placebo versus drug X at lowest dose A versus drug X at second lowest dose B versus reference drug Y (80 subjects)

Once safety of doses A and B has been demonstrated, one could proceed to:
Tier 2: Placebo versus drug X at second highest dose C versus drug X at highest dose D versus reference drug Y (80 subjects)

With the safety of drug X at doses C and D demonstrated, the study could continue with randomization of the final 160 subjects among the six groups such that there would be an equal numbers of subjects in each group at the end of the study recruitment phase. If, however, safety is shown to be an issue at the higher doses, the study would continue with only those doses judged to be safe.
II.2.E. Planned Sample
360 subjects for 6 groups with 60 subjects per group to provide 90% power (alpha = 0.05), to
a. conclude non-inferiority of the investigational drug when there is no true treatment difference
between the investigational drug X and the reference drug Y,
b. detect a true difference between the reference drug Y or the investigational drug X and placebo
under the null hypothesis, and
c. have sufficient power to examine a dose-response relationship of the investigational drug at different
doses.

II.2.F. Study Population
Approximately 400 subjects with AF or AFl of duration longer than 6 hours but less than 60 days will be
randomized. A total of 360 would be expected to complete the protocol.

II.2.G. Inclusion Criteria:
a. Either sex, between 18 and 85 years of age.
b. AF and/or AFl lasting from 6 hours to 60 days (episodes of AF/AFl of shorter duration are likely to
convert to SR on placebo while episodes of very long duration are less likely to respond to drug
therapy).
c. Anticoagulation as recommended by current therapeutic guidelines (all patients are heparinized prior
to administration of study drug and anticoagulated for at least one month after conversion to SR;
further, if AF/AFl has been present for > 48 hours, adequate anticoagulation (INR 2-3) must have
been maintained for at least 3-4 weeks prior to entry into the study or intra-atrial clot must have been
excluded by transesophageal echocardiography immediately prior to study drug administration).
d. Calculated average (beats/12 sec x 5) resting ventricular response rate < 120 bpm achieved by
administration of rate-control drugs (β-blockers and/or calcium channel blockers and/or digoxin).

II.2.H. Exclusion Criteria
a. Females who are currently pregnant or breast feeding.
b. Hemodynamic instability (CHF, hypotension, angina).
c. Resting ventricular response rate < 70 bpm in the absence of rate-control drugs or RR interval > 3 sec;
   if a patient is on rate-control drugs, these must be adjusted so that resting heart rate is not < 70 bpm.
d. QRS interval > 180 msec or QTc interval > 440 msec (in the presence of bundle branch block > 500
   msec) calculated on the average of 3 QT and RR intervals.
e. Wolff-Parkinson-White syndrome which has not undergone curative ablation.
f. History or clinical signs of thyrotoxicosis, confirmed by laboratory studies.
g. AF or AFl from reversible non-cardiac diseases (eg., pneumonia) or from acute drug effect (eg.,
excess caffeine, alcohol, bronchodilator therapy).
h. History of cardiac surgery, MI, or unstable angina within the last 3 months.
i. History of aborted SD, unexplained syncope, monomorphic or polymorphic VT.
j. Family history of prolonged QT syndrome.
k. Known sick sinus syndrome or atrioventricular block greater than first degree.
l. Presence of cardiac pacemaker or defibrillator.
m. Diastolic blood pressure (BP) > 105 mm Hg or systolic BP < 90 mm Hg.
n. Major hematological, pulmonary (necessitating continuous oxygen therapy), hepatic or renal disease
   (eg. in the case of principal route of elimination via the kidneys of investigational drug X, calculated
   creatinine clearance (ClCr) < 20 ml/min).
o. Plasma potassium level < 4.0 or > 5.5 mEq, or plasma magnesium level < 0.75 or > 1.25 mEq.
   Plasma potassium and magnesium may be corrected prior to study entry.
p. Amiodarone treatment within previous 3 months.
q. Any Class I or III antiarrhythmic agent, tricyclic or tetracyclic antidepressant, anticonvulsant, or
   phenothiazine or any other drug known to prolong the QTc within 5 half-lives before study entry.
r. Use of an experimental drug within the preceding 4 weeks.
s. Prior utilization of reference drug Y for I.V. conversion of AF/AFl.
t. History of substance abuse or dependency or ongoing psychosis.

II.2.I. Tools for assessing endpoints

Assessment Methods

Initial screening consists of a relevant medical history, physical examination, 12-lead ECG and blood sample for routine laboratory tests. An echocardiogram is obtained once the ventricular response rate has been reduced to < 120 bpm prior to study drug administration to assess LV function, valvulopathy, atrial size, and, if indicated, the presence of intra-atrial clot. A Holter recording of 3 ECG leads as well as continuous ECG monitoring (single lead) is started. Study drug is given I.V. following baseline recording for one hour.

Measure of Endpoints

ECG monitoring is continued for a maximum of 24 hours to document conversion to SR. BP and 12-lead ECGs are obtained at the start and end of the baseline period, at the time of drug administration, at 5, 10, 15, 20, 30, 60, 90, 120, 180 min, at 4, 6 and 12 hours, at the predetermined time T following study drug administration, and, finally, at the end of the study period (24 hours). A single-lead ECG may be substituted at the 5-minute intervals if obtaining a 12-lead ECG is technically impossible, but care should be taken to use a lead with an easily measurable QT interval. Preferably, the lead with the longest QT interval as determined from the 12-lead ECG should be chosen. Pharmacokinetic evaluation can be determined from venous blood samples drawn from an indwelling I.V. catheter prior to infusion of study drug (baseline) and at the times specified above.

II.2.J. Specific criteria for early withdrawal or non-utilization of subject data in analysis

a. Adverse event during drug infusion requiring stopping drug infusion; data from this subject would not be used in efficacy analysis but would still be reported in the tabulation of adverse events.
b. Protocol violation.
c. Administrative reasons.

II.2.K. Data Analysis Methods

The proportion of subjects in each treatment group achieving conversion to SR within time T can be compared using a one-way logistic model with treatment as the factor. The survival function for time to conversion can be estimated by Kaplan-Meier analysis. The following pairwise comparisons should be performed:

a. Investigational vs. reference drug: a non-inferiority test and, if positive, a superiority test of investigational drug doses A, B, C and D vs reference drug Y.

The statistical test for each of the pairwise comparisons can be made at a two-sided 0.05 statistical significance level. The predictive value of variables at baseline (e.g., duration of arrhythmia, atrial size, AF vs AFl) on the probability of conversion to SR can be assessed by multiple logistic regression analysis. Mean changes from baseline over time for mean BP, heart rate, and QRS and QTc intervals can be assessed in non-responders by analysis of covariance.

II.2.L. Comments

If safety and efficacy are demonstrated in such a study, other Phase II studies can be considered. If safety is demonstrated but there is little efficacy of I.V. drug, this does not exclude further clinical studies of drug X. Oral drug which is converted to one or more active metabolites may be effective when single dose I.V. administration is not. Further, an antiarrhythmic drug may not be useful in achieving conversion to SR but may be very effective in preventing recurrence of arrhythmia. However, if safety is clearly an issue,
especially if observed at fairly low doses and plasma levels of drug, there will be little impetus to pursuing further drug testing.

Once I.V. dosing which is both effective and safe has been established for investigational drug X, other questions best answered by physician driven Phase III studies can begin. For example, in patients failing to convert to SR, is electrical cardioversion following I.V. infusion of drug X more effective (higher percentage conversion rate or lower defibrillation threshold) than when performed in the presence of placebo or of reference drug? Further, is I.V. pre-administration of investigational drug X more effective than placebo or reference drug Y in preventing early recurrence of AF or AFl following successful cardioversion?

II.3. Long term studies

II.3.i. Conversion and maintenance sinus rhythm in patients with persistent atrial fibrillation or atrial flutter
This trial is a multicenter, randomized, double-blind, placebo-controlled parallel group study of the efficacy and safety of oral investigational antiarrhythmic drug X versus reference drug Y in converting to and maintaining SR in patients with persistent AF or AFl.

As safety has become such an issue for drugs with Class III effects, long-term Phase II studies should be initiated in a monitored setting in hospital for at least the duration of time necessary to achieve steady-state. Such an approach has been taken with both dofetilide (15) and azimilide (16) in their supraventricular Phase II programs. Also, evaluation of the efficacy of investigational drug X in the context of a long-term oral dosing protocol may now need to be done separately in patients with cardiac pathology and EF < 35 % as compared to those with better preserved EF or with normal hearts. In the former group, defibrillator implantation for SD prevention is now indicated (12,19) especially since the recent presentation of the Sudden Cardiac Death – Heart Failure Trial (12). The dual-chamber pacing capabilities of defibrillators along with various AF prevention pacing algorithms now available may significantly influence drug efficacy and safety. Further, their diagnostic abilities and memory capacity allow documentation of recurrence of atrial arrhythmia as well as of possible ventricular proarrhythmia. The following protocol therefore addresses the evaluation of investigational drug X in converting AF to and maintaining SR only in patients with normal hearts or fairly well-preserved EF (> 35%) despite cardiac pathology.

II.3.i.A. Objectives

Primary Objectives
a. To determine the efficacy of oral administration of investigational drug X at the highest tolerated dose compared to placebo and to reference drug Y in achieving conversion of persistent AF or AFl to SR within 5 half lives of the start of drug dosing (or within the time to reach maintenance plasma levels if loading doses are used).
b. To determine the efficacy of chronic oral administration of investigational drug X at the highest tolerated dose compared to placebo and to reference drug Y in maintaining SR following electrical or chemical cardioversion.
c. To determine the safety of oral administration of the investigational drug X at the most effective dose compared to placebo in subjects with persistent AF or AFl.

Secondary Objectives
a. To determine the efficacy and safety of oral investigational drug X at doses A, B and C compared to placebo in achieving pharmacological conversion of AF or AFl to SR.
b. To assess the effect of doses A, B and C of drug X and reference drug Y on ECG measures such as sinus rate, PR, QRS, QT and QTc intervals in responders or following electrical cardioversion in non-responders as well as on the ventricular response rate in non-responders.
c. To determine the range of effective plasma levels of investigational drug X in subjects with AF or AFl and its modification by concomitant cardiac or minor to moderate hepatic or renal impairment.

II.3.i.B. Primary endpoints
a. Incidence of conversion to SR within 5 half-lives of beginning study drug or, if loading doses are used, within the time to reach maintenance plasma levels (see Study Design).
b. Time to first recurrence of AF or AFl lasting at least 24 hours following electrical or chemical conversion to SR.
c. Incidence of serious side effects.

II.3.i.C. Secondary endpoints
a. Mean time to conversion in responders.
b. Ventricular rate in non-responders with respect to baseline.

II.3.i.D. Study Design
Subjects who meet the study inclusion/exclusion criteria are admitted to telemetry in hospital for the duration of time necessary to reach the plateau effect of investigational drug X or reference drug Y whichever is longer. Randomization is done in a double-blind fashion to oral investigational drug X at doses A, B, or C, the reference drug Y or placebo. Maximum dose of investigational drug X (i.e., dose C) should not exceed that proven safe in Phase III studies of mortality. If investigational drug X is eliminated primarily by the kidneys, dosage adjustment is performed according to calculated Clcr, as follows:
   a. ClCr 40 – 60 ml/min – half of the randomized dose is given,
   b. ClCr 20 – 40 ml/min – one quarter is given.

A 12-lead ECG is obtained just before the first dose on all days while in hospital. If the QTc interval is found to increase by > 15 % over baseline, the dose is halved. If QTc interval exceeds 550 msec or increases by > 25 % over baseline, the subject is withdrawn from the study. It is important to note that in the design of such a study, reference drug Y should use the same predominant route of elimination/metabolism as the investigational drug X, as the dosage adjustments described above may penalize the reference drug Y if it does not depend on renal clearance. On the other hand, when elimination of investigational drug X does not depend on renal clearance and therefore dose adjustments are not necessary, there is a risk of significant toxicity in the case of reference drug Y when its clearance does depend on the kidneys.

In subjects who are exposed to drug for a duration of time to achieve plateau levels and have not converted spontaneously to SR, electrical cardioversion is attempted. Following conversion to SR, subjects are monitored for an additional 24 hours. Those in whom SR cannot be achieved or maintained for 24 hours following electrical or pharmacological conversion are withdrawn from the study. Subjects who are at this point in SR are discharged from hospital on the same dose of study drug as last given in hospital, or on maintenance dosing in those in which the hospitalization period was used for drug loading. Oral anticoagulation is continued for only one month in patients at low risk for cardiac emboli but throughout the course of the study in patients at moderate to high risk (6). Follow-up clinic visits are scheduled at 2, 4, 6, 8, 10, 12 weeks and then every 3 months until one of the study end points is reached: relapse to AF or AFl for at least 24 hours, documented by ECG, or maintenance of SR for one year. Twelve-month survival data (freedom from AF or AFl) is collected for all randomized subjects regardless of treatment duration.

II.3.i.E. Planned Sample
Sample size for a double-blind study can be calculated using the method of Schoenfeld (17) (proportional hazards regression model), assuming a two-sided hypothesis test of the primary endpoint at a significance level of 5 %. Total sample size will then depend on the number of groups (investigational drug X doses studied and on the presence or absence of a reference drug Y group). The distribution of time to first event documented by ECG is assumed to follow an exponential distribution. If the median time to first event is
estimated to be no more than 90 days for placebo, then a sample size of 190 subjects in each treatment group will allow a hazard ratio of 0.67 to be detected with 0.90 probability. With possibly a 10-20% drop out rate, then 210-230 subjects should be enrolled in each treatment group.

II.3.i.F. Study Population
If all the groups mentioned above are used in the study, then a minimum total sample size of 1260 subjects with AF or AFI of duration longer than 2 weeks but less than 26 weeks will be recruited. A total of about 1140 subjects would be expected to complete the protocol. If fewer doses are studied or if no appropriate reference drug Y is available for comparison, then the total population size is correspondingly reduced.

II.3.i.G. Inclusion Criteria
a. Either sex between 18 and 85 years of age.
b. AF and/or AFI lasting from 2 to 26 weeks, confirmed by ECG.
c. Anticoagulation in all patients with a therapeutic INR of 2-3 for at least 3-4 weeks prior to beginning study drug, or, if a patient is not currently anti-coagulated, heparin I.V. is administered and oral anticoagulation is begun. Study drug can be administered early on I.V. heparin if transesophageal echocardiography shows no intra-atrial thrombus.
d. Average resting ventricular response rate < 120 bpm (see Section II.2.G).

II.3.i.H. Exclusion Criteria
Same as listed in section II.2.H with the addition of the following:
a. Females who plan to become pregnant during the course of the study or, if sexually active, are not using a hormonal contraceptive as well as a vaginal spermicide.
b. Presence of severe valvulopathy, such that surgical intervention is considered a possibility within the time course of the study.
c. Ischemic or non-ischemic CM with EF < 35%.
d. Presence of other life-threatening disease with survival expected to be < 2 years.
e. Use of digoxin, unless plasma levels remain constant.
f. Absolute contraindications to anticoagulation therapy.
g. Any unresolved drug-induced organ toxicity.
h. Concomitant therapy with drugs known to affect the metabolism or elimination of investigational drug X or reference drug Y.
i. Patients who previously in the opinion of the investigator have failed an antiarrhythmic drug of the class being tested for efficacy reasons.
j. Patients who have previously participated in a study of investigational drug X or have used reference drug Y for oral conversion to and/or the subsequent maintenance of SR.
k. If Phase III studies in subjects with underlying ischemic/non-ischemic CM and but less depressed LV function (EF 35-50%) demonstrate significantly higher mortality in the presence of investigational drug X compared to placebo, such patients should also be excluded in chronic studies of efficacy in AF or AFI.

II.3.i.I. Tools for assessing endpoints
Initial screening should include a relevant medical history, complete physical examination, 12-lead ECG, chest X-ray, echocardiogram and blood sample for routine laboratory tests as well as those indicated by the known side effect profile of investigational drug X and reference drug Y. In-hospital, a 12-lead ECG is done each day before the morning dose of study drug. If electrical cardioversion is done or if chemical conversion occurs, an ECG should be obtained immediately to document SR and also 24 hours later. Blood tests before discharge from hospital should include serum chemistry, hematology, and analysis of levels of investigational drug X, reference drug Y and principal active metabolite(s), if applicable.
At follow-up visits, vital signs, cardiopulmonary specific physical examination, event symptom severity checklist (prior to ECG), 12-lead ECG, serum chemistry, hematology, blood samples for analysis of investigational drug X, reference drug Y and active metabolite(s) and assessment of concomitant medications and adverse events are done. Assessment of QOL (SF-36) and the Brignole Atrial Fibrillation Symptom Checklist can be completed on a monthly basis. As well, drug accountability and, in females of childbearing potential, a serum pregnancy test should be performed. At the final visit, either on reaching the primary endpoint or on completion of the total study duration, in addition to the procedures described for the follow-up visits, a complete physical examination, a healthcare resource utilization questionnaire and a chest X-ray are done.

A 12-lead ECG is obtained at each visit, reviewed by the investigator and sent to a central facility which generates a report including an interpretation of the patient’s rhythm and interval calculations. If the investigator and central ECG facility interpretations do not agree with respect to possible AF, AFI or SVT, the report should be sent to an Event Committee for a final decision.

Adverse Events (AE) are defined as any undesirable clinical experience during the study, whether or not related to the study drug, including an exacerbation of a preexisting condition. A serious AE is one that results in death or is life threatening, results in hospitalization or prolongation of current hospitalization, results in persistent or significant disability/incapacity, is a congenital anomaly/birth defect or is judged to be medically significant. All AEs should be judged as to their severity (mild, moderate, severe) and as to their causality (doubtful, possible, probable).

II.3.i.J Specific criteria for early withdrawal of subjects from the study
a. ECG criteria: ventricular fibrillation, polymorphic VT of any duration, sustained monomorphic VT (> 30 sec) or symptomatic non-sustained VT, incessant VT (recurrent VT < 30 sec interrupted by a few sinus beats), QT<sub>c</sub> > 525 msec, resting SR remaining < 50 bpm after conversion to SR.

b. Pregnancy

c. Noncompliance: missing any 3 scheduled visits or any 2 serum pregnancy tests, stopping contraceptive methods or taking less than 80 % of study drug.

d. Adverse events or organ toxicity such that withdrawal of the subject is recommended.

e. Voluntary withdrawal by the subject.

f. Protocol violation.

g. Administrative reasons.

II.3.i.K. Data Analysis Methods
Primary endpoints
All intent-to-treat subjects are used in the analysis of the primary efficacy endpoints. The Cochran-Mantel-Haenszel test or the Fisher’s exact test can be used to compare the percentage of randomized subjects pharmacologically converting to SR in each investigational drug X group compared to placebo. Similarly, comparison can be made between each investigational drug X group vs. reference drug Y group. Time from the start of the efficacy period to the date of the first documented AF, AFI or SVT of > 24 hours duration is measured and displayed using Kaplan-Meier estimates of the survival curves. Treatment comparisons can be made using log-rank test, differences between the time to event distributions can be quantified by the hazard ratio from proportional hazards regression, Kaplan-Meier estimates of the median time to event, and Kaplan-Meier estimates of the proportion event-free through one year. Pairwise comparisons between investigational drug X at doses A, B or C vs reference drug Y can be performed, as well as pairwise comparisons between investigational drug X at doses A, B or C vs placebo. The predictive value of variables at baseline (e.g., duration of arrhythmia, atrial size, AF vs AFI, LV systolic dysfunction) on the probability of conversion to SR or on time to recurrence of atrial arrhythmias can be assessed by multiple logistic regression analysis.
All subjects randomized to a treatment group are used in safety summaries. Incidence of AEs is compared between all groups and frequencies are tabulated by body system, treatment duration, causality and severity. Mortality rates can be calculated for each group. Treatment comparisons can be made by calculating the mortality relative rate using a Poisson model and the mortality relative risk using the Cochran-Mantel-Haenszel method (18). Efficacy and safety data can be summarized separately for subgroups defined, for example, by sex, age, race/ethnicity, baseline cardiovascular disease state, EF, smoking history, digoxin or β-blocker use and presented as point estimates and estimated standard errors. Descriptive statistics are used to calculate the clinical laboratory data and ECG intervals which can then be presented in shift tables and/or in shift plots.

Secondary endpoints
Time to conversion can be determined by the number of hours from first dose in each group and displayed graphically by means of the Kaplan-Meier (product-limit) method. Ventricular response rates in non-converters at first ECG before study drug administration and prior to electrical cardioversion can be presented in the different groups with descriptive statistics. Change and percentage change can be compared between placebo, investigational drug X groups and reference drug Y group by analysis of variance and Bonferroni t-test. QOL comparisons can be made using analysis of variance or if assumptions of normality and equal variances are not valid, a non-parametric analysis of covariance can be used. Symptom frequency load during the first recurrence can be assessed by constructing contingency tables of number of treatment groups versus number of pre-specified symptoms reported during the first recurrence of atrial arrhythmia and analyzed with the appropriate chi-square test.

II.3.i.L. Comments
Demonstration of efficacy and safety in such a study suggests the possibility of beginning drug on an outpatient basis. However, if the dose must be modified during the first few days either because of excessive effects on the ECG or because of significant proarrhythmic effects, further Phase II studies will require the same initiation protocol, and, once brought to market, initiation in a monitored hospital setting will be mandatory. This would significantly limit this drug’s use appropriately to specialists trained in its administration and, perhaps inappropriately, for those for whom financial restrictions and hospital bed availability are not an issue.

Further Phase III studies could determine if defibrillation thresholds for AF are decreased by oral pretreatment with investigational drug X versus placebo or reference drug Y, or in patients with pacemakers or pacemaker/defibrillators with programming capabilities for pacing prevention of AF and AFl, it could be determined if such programs are more effective in preventing recurrence of these arrhythmias in the presence of drug X versus placebo or reference drug Y.

II.3.ii. Prophylactic treatment of paroxismal atrial fibrillation or atrial flutter
The trial is a multicenter, randomized, double-blind, placebo-controlled, parallel-group clinical study to assess the efficacy and safety of the oral investigational antiarrhythmic drug X versus reference drug Y in the prophylactic treatment of paroxysmal atrial fibrillation and atrial flutter.

With the accumulation of data from studies on the acute conversion of AF and AFl with I.V. or oral investigational drug X and the prevention of recurrence of arrhythmia following pharmacological or electrical cardioversion, it is usually possible to identify a dose range which is effective and tolerated. The number of investigational drug groups can now be limited to one or two doses for evaluating efficacy and safety when used for other indications, such as the prevention of paroxysmal AF (PAF). Accumulating data from mortality studies in subjects with moderate systolic LV dysfunction in ischemic or non-ischemic CM will determine whether such a study should be limited to subjects with PAF or PAFl and normal hearts or whether two strata can be studied as described below. If so, as in section II.2, such a chronic study should
exclude patients with ischemic or non-ischemic CM and LVEF < 35% for whom defibrillator implantation is indicated.

II.3.ii.A. Objectives

Primary Objectives

a. To assess the efficacy of oral investigational drug X at doses A and B compared to placebo and reference drug Y in prolonging the arrhythmia-free period, i.e., time from first dose of study medication to first ECG-documented symptomatic AF, AFI or SVT.

b. This primary objective can be assessed separately in two strata:
   - subjects with ischemic heart disease (IHD) and/or CHF (EF >35%)
   - subjects with neither of the above

Secondary Objectives

The following secondary objectives can be determined according to stratification between presence or absence of CHF and IHD.

a. To assess the efficacy of investigational drug X at doses A and B versus placebo and versus reference drug Y in reducing the total number of symptoms reported during the first symptomatic event.

b. To assess the efficacy of investigational drug X at doses A and B versus placebo and versus reference drug Y in reducing the frequency of symptomatic events.

c. To assess the efficacy of investigational drug X at doses A and B versus placebo and versus reference drug Y in reducing the total supraventricular arrhythmia burden in patients (i.e., frequency, duration and severity).

d. To assess the effect of investigational drug X at doses A and B versus placebo and versus reference drug Y on patient QOL.

e. To assess the number of days in-hospital and number of emergency room visits due to symptomatic PAF, PAFI or PSVT (following initial discharge for subjects hospitalized for drug loading or initiation) between groups receiving investigational drug X at doses A and B versus placebo and versus reference drug Y.

f. To assess the efficacy of investigational drug X at doses A and B versus placebo versus reference drug Y in reducing the number of asymptomatic events.

And, in both strata combined:

g. To assess the safety of investigational drug X at doses A and B versus placebo and versus reference drug Y according to incidence and severity of AEs, ECG changes, and new laboratory and chest X-ray abnormalities.

II.3.ii.B. Study Endpoint

Time to a SR containing day after the second confirmed occurrence of symptomatic PAF, PAFI or PSVT.

II.3.ii.C. Study Design

Subjects who meet the study screening inclusion/exclusion criteria are instructed in the use of a TTM. SR at the beginning of the screening period is documented during a training call. Over the following month, subjects transmit a recording obtained whenever they have symptoms suggestive of an arrhythmia. Once symptomatic AF or AFI is documented, subjects transmit again when SR returns. Randomization then occurs between investigational drug X at dose A or B, placebo or reference drug Y. If no spontaneously terminating episode of AF or AFI is documented during a one-month screening phase off antiarrhythmic drugs, the subject is withdrawn from the study. Study drug is initiated only once other Class I or III drugs are stopped for at least 5 half-lives.
Initiation of study drug using in-patient hospitalization, continuous monitoring and a 12-lead ECG prior to each morning dose versus close out-patient monitoring and daily transmission of a rhythm strip prior to morning dose will depend on the incidence of proarrhythmia observed in prior trials with investigational drug X in the absence and the presence of known IHD or CHF as well as on the clinical experience with reference drug Y. In the case of significant renal clearance of drug as well as risk of torsade de pointes, in-hospital initiation of study drug using the dosing design described in section II.2.D should be used. When clearance is primarily by hepatic metabolism and there is little risk of torsade de pointes, out-patient initiation of study drug can be considered, at least in cases of PAF or PAFI in normal hearts.

During the maintenance period of study drug administration, a TTM recording should be transmitted whenever the subject experiences symptoms suggestive of arrhythmia. An Event Symptom Severity Checklist should be obtained by the central TTM facility before transmission. If the subject is in a location where an ECG can be obtained, this checklist should be administered by health care personnel prior to recording the rhythm. All tracings should be sent to the central facility. Documentation of return to SR can be made by TTM, ECG or telemetry. If return to SR does not occur spontaneously, then Class II and/or Class IV drugs and/or digoxin can be administered to slow the ventricular response rate to < 100 bpm, and, if arrhythmia persists, electrical cardioversion can be done. Neither Class I nor Class III drugs should be used. Subjects then continue on the same study drug assigned until documentation of return to SR after the second confirmed symptomatic atrial arrhythmia occurs or until end of study at 12 months.

II.3.ii.D. Planned Sample
Sample size is calculated as in section II.2.E using Schoenfeld’s method, assuming two-sided tests, and a time to first event that follows an exponential distribution with a median of 90 days in the placebo group. Since the assigned alpha applies only to the time to first symptomatic event, Hochberg’s procedure is applied to the sequence of the 2 tests (17). Allowing for drop-outs, approximately 220 subjects are projected for each group. This would allow an investigational drug to placebo hazard ratio of 0.67 to be detected with 0.90 probability at an alpha of 0.045.

II.3.ii.E. Study Population
If the investigational drug is administered at doses A and B, and both placebo and reference drug Y are used, then a study sample size of 880 subjects with PAF or PAFI are required in each stratum after successful screening. It should be pointed out that the logistics of screening a large patient population to document a symptomatic arrhythmia episode within a one-month period may significantly dampen enthusiasm for a multi-group trial. To be practical and finish the study recruitment in a timely manner, only the most promising dose of investigational drug X may be compared to placebo. Comparison to a drug already on the market may have to wait for positive results of a simpler investigational drug X versus placebo trial.

II.3.ii.F. Inclusion Criteria
Screening Period
a. Either sex between 18 and 85 years of age.
b. Symptomatic AF within the previous 6 months, documented by ECG, TTM, or rhythm strip but not by Holter monitor.
c. Symptoms severe enough to significantly interfere with the subject’s usual activities.
d. SR at the time of starting the screening period.

For Randomization
a. Within 30 days from starting screening, at least 1 episode of symptomatic AF or AFI documented by TTM or ECG (if no episode occurs on antiarrhythmic drugs, at least 1 episode during a second 30-day screening period after discontinuing these medications).
b. SR immediately prior to starting study drug.
II.3.ii.G. Exclusion Criteria

Screening Period

a. Women currently pregnant or breast-feeding, or who plan to become pregnant during the course of the study or are unwilling to use hormone contraceptives with a vaginal spermicide; women postmenopausal for less than one year or not surgically sterile (tubal ligation at least 3 months prior to screening) unless they have been using hormonal contraceptives for at least 3 months prior to entry into the screening phase.

b. AF or AFl due to acute electrolyte disturbance, hyperthyroidism, pericarditis, or any other acute reversible illness.

c. Electrical cardioversion within 60 days prior to screening.

d. Syncope, angina pectoris or pulmonary oedema that is precipitated by attacks of arrhythmia (ventricular or supraventricular).

e. Polymorphic VT of any duration, sustained (> 30 sec) monomorphic VT, aborted SD or undiagnosed syncope.

f. Wolff-Parkinson-White syndrome that has not undergone curative ablation.

g. QRS interval > 180 msec or QT interval > 440 msec (in the presence of bundle branch block > 500 msec) or a family history of prolonged QT syndrome.

h. Presence of an implanted pacemaker or defibrillator.

i. History of Class IV New York Heart Association CHF or current acute decompensated CHF. Class I – III patients must have been stable for at least one month.

j. Ischemic/non-ischemic CM with EF < 35%.

k. Cardiac surgery, MI, or unstable angina within the last 3 months.

l. Symptomatic severe valvular disease for which surgery is considered within the duration of the study.

m. Stroke or a reversible ischemic neurologic deficit within the last 3 months.

n. Diastolic BP > 105 mm Hg (acceptable value must be present for at least one month on antihypertensive therapy) or systolic BP < 90 mm Hg.

o. Major hematological, pulmonary (requiring continuous oxygen therapy), hepatic or renal disease (calculated ClCr < 20 ml/min).

p. Unresolved drug-induced organ toxicity.

q. Substance abuse or dependency or ongoing psychosis.

r. Previous failure of efficacy of the class of antiarrhythmic drug being tested.

s. Prior participation in a randomized trial of investigational drug X.

t. Prior exposure to reference drug Y.

u. Any investigational drug use in the 30 days before screening.

For Randomization

a. Electrical cardioversion during the screening period.

b. Resting SR below 50 bpm.

c. QTc interval > 440 msec, calculated as the average of 3 QT and RR intervals recorded during SR.

d. Class I or Class III drugs within 5 half-lives or amiodarone within one month prior to receiving the first dose of study drug or during the study.

e. Use of drugs which prolong the QTc interval when the investigational drug X or reference drug Y is a Class III drug and of any drug that may potentiate a known serious side effect of investigational drug X or reference drug Y for at least 5 half-lives prior to receiving the first dose of study drug or any time during the study.

f. Potassium < 4 mEq/L or > 5.5 mEq/L or magnesium below the lower limit of normal; both of these electrolyte levels may be corrected prior to beginning study drug.

g. Hepatic dysfunction (ALT, AST > 2 x normal), severe renal dysfunction (calculated ClCr < 20 ml/min).
II.3.i.H. Tools for assessing endpoints
A baseline visit for randomized subjects is performed within 7 days prior to first dose of study drug. This includes a medical history, complete physical examination, chest X-ray, hematology, blood chemistry (electrolytes including magnesium, renal and hepatic function tests and others as suggested by the side effect profile of both investigational drug X and reference drug Y), blank sample for drug and metabolite plasma levels and serum pregnancy test for women of child-bearing potential. An echocardiogram should have been done to determine eligibility for the study prior to screening. Potassium and magnesium levels must at all times be maintained in the normal range. Immediately prior to the first dose of study drug, vital signs are taken, an event severity checklist is done and then a 12-lead ECG is taken to confirm SR and QTc < 440 msec and serum chemistry is repeated stat if baseline chemistry is not done on day 1 of dosing. Finally, QOL is assessed by SF-36 and Brignole Atrial Fibrillation Symptom Checklist.

Clinic visits and TTM are scheduled often to document asymptomatic/minimally symptomatic recurrence of atrial arrhythmia. TTM transmission can be done weekly if no clinic visit is scheduled for that week with Event Symptom Severity Checklist done before transmission. Clinic visits can be scheduled at the end of week 2, 4, 8, 12, 16, 24, 36, and 52. At each visit the following are recorded: vital signs, Event Symptom Severity Checklist before a 12-lead ECG, blood chemistry, hematology, blood levels of investigational drug X, reference drug Y and possible metabolites, assessment of adverse events and verification of concomitant medications. Also, drug accountability is done at the beginning of the maintenance phase and both drug accountability and QOL assessment are done at weeks 4, 8, 12 and 24 and 52. A pregnancy test is done at monthly intervals in women of child-bearing potential following the baseline test until the end of the patient’s participation in the study. At the last visit (week 52 or on early completion or withdrawal) a complete physical examination is performed and a chest X-ray should be obtained.

Post-treatment follow-up is done at day 7 and day 30 following study termination. The same data is collected as above as well as a pregnancy test if indicated at day 30 post-treatment. During this 30-day period subjects continue to transmit recordings during symptoms suggestive of arrhythmia. The TTM is returned only at the end of this 30 day period.

Twelve-lead ECGs obtained at each scheduled or unscheduled visit and any TTM or telemetry recordings are reviewed by the investigator. All the above are sent to a central facility which will generate a report of rhythm diagnosis and interval measurements. If there is disagreement with respect to the rhythm diagnosis, the tracing should be submitted to an Event Committee. QOL questionnaires will be completed at home prior to the appropriate visit. Safety (AEs) will be assessed as described in section II.3.K.

II.3.i.I. Data Analysis Methods
Primary endpoints
The log-rank test can be used to compare the distribution of the time to first symptomatic event (from day of randomization) between patients in the various groups. Hochberg’s procedure can be applied to the sequence of the 2 tests (17). Differences between the time to event distributions can also be quantified by the hazard ratio from a proportional hazards regression, Kaplan-Meier estimates of the median time to event, and Kaplan-Meier estimates of the proportion event-free through 6 months or 12 months. The presence or absence of Class II/III CHF can be used as a stratification variable for the log-rank test and proportional hazards regression in the CHF/IHD stratum.

Secondary endpoints
The number of 6 pre-specified symptoms reported during the first confirmed event (Event Symptom Severity Checklist) are totaled. The total count can be compared between treatment groups, using a chi-square test with rank scores. Also, the percentage of subjects with each specific symptom can be compared using a chi-square test. Symptomatic event rates can be compared between treatment groups and analyzed using a Poisson regression model (18). ECG data can be used to assess subject arrhythmia status over time
and this latter can be combined with symptom count and severity data to classify each subject into an ‘SVA burden state’. The proportion of subjects in each state or combination of states can be compared between treatment groups using a logistic regression or hazard regression model.

The Medical Outcome Study Short Form (SF-36) Health Survey consists of 36 questions which assesses 8 areas one of which is physical functioning. This subscale can be normalized to a 0 – 100 % scale from worst to best functioning. Descriptive statistics (n, mean, median, minimum, maximum and standard errors) can be used to summarize the data and analyses of variance can be used to test for treatment differences in physical functioning. Resource utilization can be assessed by evaluating the number of days in-hospital after in-hospital initiation of study drug or the number of emergency room visits due to atrial arrhythmias normalized by the number of days at risk. This can be modeled using Poisson regression and the relative risk and standard error in investigational drug X group versus placebo or versus reference drug Y can be estimated.

Other analyses can compare the rate of asymptomatic arrhythmias between treatment groups using a Poisson regression model. Other SF-36 QOL measures (eg., general health, vitality, social functioning and mental health) can also be assessed as described above. The Brignole Atrial Fibrillation Symptom Checklist quantifies the presence of 5 symptoms which are disease specific (palpitations, shortness of breath at rest, shortness of breath during physical activity, fatigue during mild physical activity and fatigue at rest) during the past 4 weeks by means of a score scale (0- absent to 10- present ). This too can be summarized with descriptive statistics or frequency tables at the scheduled time points. Analysis of variance can be used to test for treatment differences using factors such as baseline value, sex or baseline EF as covariates.

The incidence of AEs can be summarized using tables. Descriptive statistics can be used to summarize laboratory data as well as ECG values (QT, QTc, PR, QRS and heart rate during sinus rhythm or during atrial arrhythmia) by visit, including absolute values, change and percentage change from baseline. The latter can be presented in shift tables or plots. Reason for drug discontinuation should be presented in frequency tables. Compliance (number of tablets taken divided by number expected to be taken) can be summarized with descriptive statistics.

II.3.ii.J. Specific criteria for early withdrawal or non-utilization of subject data in analysis

a. ECG criteria: ventricular fibrillation, sustained monomorphic VT (> 30 sec), incessant VT (recurrent VT with episodes lasting < 30 sec and interrupted by a few sinus beats), polymorphic VT or a QTc > 525 msec on 12-lead ECG or rhythm strip. If a patient is withdrawn from the study for ECG criteria, a blood sample should be obtained for plasma level of investigational drug X and reference drug Y as well as any possible active metabolites and for electrolytes (including Mg++).

b. Pregnancy.

c. Noncompliance: missing any 3 or 2 consecutive clinic visits, a total of 3 or any 2 consecutive TTM transmissions, or 2 serum pregnancy tests; stopping hormonal contraceptives or vaginal spermicide; use of any excluded concomitant medication.

d. Adverse events or organ toxicity that require subject withdrawal from the study.

e. Protocol violation.

f. Administrative reasons.

II.3.ii.K. Comments

The above Phase II studies represent the core of investigational drug development for the indication of AF or AFl. They provide an excellent evaluation of drug efficacy and safety of both I.V. and oral forms at various doses compared to placebo. Although to date not often done within the same study, incorporating the most effective reference drug Y also provides useful information. It is such a comparison which in
fact determines whether investigational drug X is eventually marketed just as a useful alternative to presently used drugs or whether it can in fact be considered the drug of first choice.

III. PHASE III STUDIES IN PATIENTS AT RISK OF SUDDEN DEATH DUE TO VENTRICULAR ARRRHYTHMIAS

The Cardiac Arrhythmia Suppression Trial (1,2), a Phase III study of safety and efficacy of flecainide and encainide in a large patient population at risk of SD in the chronic phase of MI, was carried out well after the introduction of these two drugs to market. Since the finding in this study that mortality was increased on drug, such Phase III testing has been more recently done early and, in fact, prior to obtaining regulatory approval. Indeed, observations made during such a study often can help in designing safer Phase II studies. For example, the most recently published mortality trial compared azimilide to placebo in patients with recent MI (6-42 days) and who were at moderate to high risk for SD as predicted by poor LV systolic function (EF < 35%) with or without low heart rate variability (20). No difference in overall mortality was observed on drug. Although azimilide would therefore not be indicated for prevention of SD, such a study does confirm the safety of its use for atrial arrhythmias in this population and permits inclusion of such patients in clinical trials of drug for supraventricular arrhythmias.

It is unlikely, however, that such placebo-controlled mortality studies in moderate to high risk populations will continue to be performed. The recent publication of MADIT II (19) and the presentation of the SCD-HeFT results (12) have shown the superiority of implantable defibrillators versus placebo in patients with poor LV function and IHD as well as versus placebo and amiodarone in non-ischemic CM (12). The indication now exists, therefore, for implantation of a defibrillator in such patients. As a consequence, long-term mortality trials of oral investigational drug X versus placebo are, in our opinion, no longer ethical in this population. Instead, the focus will likely be shifted to that group of patients with ischemic or non-ischemic CM and with only moderately or minimally depressed LV function (EF 36-50%). These patients, while less at risk for SD from ventricular arrhythmias as a result of the underlying cardiac pathology, may however be at risk for greater mortality on investigational drug X if there is any proarrhythmic potential of the new drug. Such a suggested study is presented.

III.1. Outline of a typical development plan

The study should be a double-blind, placebo-controlled, parallel design to determine the effect of two doses of orally administered investigational drug X versus placebo on survival in ischemic or non-ischemic cardiomyopathy and low to moderate risk of sudden death.

III.2. Long-term studies

III.2.A. Objectives

Primary Objective
To evaluate the effects of investigational drug X at dose A versus placebo or dose B versus placebo on all-cause mortality based on intention-to-treat in subjects at low to moderate risk of SD, i.e., with either ischemic or non-ischemic CM and LV EF 36-50%.

Secondary Objectives
To evaluate the effects of investigational drug X at doses A and B combined versus placebo on all-cause mortality, based on intention-to-treat analysis.

Tertiary Objectives
These analyses will be based on ‘on-treatment’ observations, i.e up to 30 days following discontinuation of the study drug.
a. To evaluate the effects of investigational drug X at dose A versus placebo or dose B versus placebo on all-cause mortality.

b. To evaluate the effects of investigational drug X at doses A and B combined on arrhythmic, cardiac, and non-cardiac mortality.

c. To determine the effect of beta blocker and angiotensin converting enzyme inhibitor use on all-cause mortality.

III.2.B. Study Endpoint

The study endpoint is death due to any cause.

III.2.C. Study Design

Patients who meet the study inclusion/exclusion criteria are enrolled in the study and first undergo a 24-hour Holter to determine heart rate variability (HRV) which is analyzed at a central facility. The baseline visit (within 48 prior to initiation of study drug) also includes a complete physical examination, review of concomitant medications, 12-lead ECG, stat potassium and pregnancy test (if applicable) and chemistry and hematology measurements.

The study subjects are then equally randomized among the treatment groups. Short-term hospitalisation during initiation of treatment to achieve plateau drug levels and effect on ECG intervals will depend on the need for continuous monitoring because of risk of early proarrhythmia as suggested by early I.V. or oral single-dose Phase II studies. Dosing for investigational drug X whose elimination is highly dependent on renal function will be adjusted as described in section II.3.D. Further adjustment of drug dose in the case of a Class III drug will also be influenced by the effect on QTc interval as described in the same section.

Total study duration is two years. Follow-up visits can occur at 2 weeks, and 1, 4, 8, 12, 16, 20 and 24 months and 1 month following the end of the study or 4 weeks following withdrawal from the study. The same evaluations as at baseline, except for the Holter recording, are done at each of these visits. As well, drug compliance evaluation can be performed at months 1, 4, 8, 12, 16, 20 and 24. Serum pregnancy test should be done monthly in patients of childbearing potential. Patients who withdraw early from the study should be followed by telephone contact until a time corresponding to two years since study drug initiation (for assessment of outcome on an intention-to-treat basis).

III.2.D. Planned Sample

If it is assumed that the all-cause mortality in such low to moderate risk patients is 5-8% over two years, and investigational drug X will reduce this rate by at least 45%, then using the method of Schoenfeld (17), to ensure 90% power at a significance level of 0.04, at least 4000 subjects must be recruited to show significantly reduced mortality on investigational drug X. If however, the purpose of the study is to only demonstrate safety of drug in this population, then a hypothesis of non-inferiority would require a smaller population sample size.

III.2.E. Study Population

A total of approximately 4000 subjects with ischemic or non-ischemic CM and LV EF of 36 - 50% will be enrolled.

III.2.F. Inclusion Criteria

a. Ischemic or non-ischemic CM.

b. LV EF (36-50%).

c. Male or female between the ages of 18 and 75 years.

d. If female, post-menopausal for more than one year, surgically sterile (at least 3 months post tubal ligation) or on oral contraceptives for at least 3 months prior to study entry which they will continue to use in addition to a vaginal spermicide during the study.
III.2.G. Exclusion Criteria

a. If female, currently pregnant or breast feeding, or plan to become pregnant during the course of the study.
b. History of torsade de pointes or any form of polymorphic VT.
c. History of sustained (> 30 sec) monomorphic VT.
d. Syncope or aborted SD.
e. Resting heart rate below 50 bpm.
f. Second-degree (Mobitz II) or third degree AV block without a permanent pacemaker.
g. Implantable cardiac defibrillator.
h. QTc interval measuring 450 msec or greater at enrollment or a family history of long QT.
i. Wolff-Parkinson-White syndrome not having undergone curative ablation.
j. Decompensated CHF at the time of enrollment.
k. Unstable angina pectoris.
l. Angioplasty or coronary artery bypass grafting within one month prior to study entry.
m. Severe symptomatic valvular disease for which surgical intervention is considered within the time course of the study.
n. Stroke with significant neurological deficit.
o. Uncontrolled hypertension (systolic BP > 170 mmHg or diastolic BP > 100 mmHg) at enrollment. BP must be controlled adequately for one month prior to study entry.
p. Known concurrent illness likely to affect survival within 2 years of study entry.
q. History of chronic liver disease or significant kidney disease (CrCl < 20 ml/min).
r. History of unresolved organ toxicity secondary to drug use.
s. Prior to randomization, baseline potassium < 4 mEq/L or > 5.5 mEq/L. Potassium level can be corrected and must be maintained within the acceptable range during the study.
t. Amiodarone use within one month prior to enrollment; current use of Class I or III antiarrhythmic drugs or within 5 half-lives prior to beginning study drug.
u. Use of drugs which prolong QTc interval during or within 5 half-lives prior to study.
v. Use of any non-approved investigational drug within 30 days prior to enrollment.
w. Previous participation in a trial of investigational drug X.
x. Known alcohol abuse or illicit drug use or current, diagnosed psychosis.
y. Unwillingness or inability to give written informed consent.

III.2.H. Tools for assessing endpoints

Measure of LV EF can be performed by nuclear isotope or invasive contrast ventriculography, or by 2D-echocardiography. Low HRV on 24-hour Holter is assessed by standard methods and defined as less than 20 U. A 12-lead ECG to assess study drug effects on QTc interval is performed prior to and daily during study drug initiation if indicated with in-hospital monitoring for dose adjustment, and at each follow-up visit.

All-cause mortality is documented, but is further subclassified according to the following definitions:

a. Cardiac mortality: all deaths except those due to a demonstrated non-cardiac cause.
b. Non-cardiac mortality: death due to demonstrated non-cardiac cause including vascular mortality (stroke, embolism, ruptured aneurysm).
c. Arrhythmic mortality: - death within 1 hour on the onset of new symptoms in the absence of severe LV dysfunction or shock; unwitnessed death in an apparently stable patient; unresuscitated ventricular fibrillation; sudden or non-sudden cardiac death with documented or suspected arrhythmia; unwitnessed, sudden, presumed cardiac death.

Adverse events are defined as in section II.3.K.
III.2.I. Data Analysis Methods

The primary efficacy analysis is performed on all-cause mortality in all randomized patients. Kaplan-Meier curves are used to estimate survival for each treatment group. Median time-to-event, with 95% confidence intervals, are determined for each group. More complex statistical analyses can also be performed for each active to placebo comparison. Secondary efficacy analyses based on intention-to-treat observations are performed as for the primary efficacy analysis. The same statistical methods are used for the on-treatment efficacy analyses but data from patients who withdraw early from the study is gathered only until one month following discontinuation of study drug. Interim efficacy analyses can be conducted at pre-determined times during the course of the study and should be based on all-cause mortality. Both, statistically significant decrease or increase in mortality on investigational drug X will be assessed but the study should only be prematurely terminated for a statistically significant increase in all-cause mortality on investigational drug X.

Safety data can be presented using graphs and tables of summary statistics by treatment group. Frequency of AEs, AEs per unit time of study drug exposure are shown. Laboratory data (chemistry, hematology and ECG measurements) are shown as box plots over time. Further, ECG data can be summarized as actual values, change from baseline and percent change from baseline. Efficacy and safety data can be summarized separately for such subgroups as sex, age, ischemic versus non-ischemic CM, Class I versus Class II/III CHF or utilization of β-blockers and/or angiotensin converting enzyme inhibitors.

Interim analyses should be done at predetermined intervals by a statistician not associated with the study and are based on the primary end-point of all-cause mortality. Both, statistically significant improvement or decrease in mortality on investigational drug X should be noted, but recommendation to prematurely terminate the study should only be made if decreased survival is observed on drug X.

III.2.J. Specific criteria for early withdrawal or non-utilization of subject data in analysis

Early withdrawal of subjects from the study is based on the same criteria as in section II.3.J however these data are still used in the intention-to-treat efficacy analyses. Subject data only is not used in the case of protocol violation or for administrative reasons.

III.2.K. Comments

The demonstration of increased mortality in the presence of investigational drug X in patients with underlying cardiac pathology and mild to moderate LV dysfunction would essentially limit the use of the new drug to patients with normal hearts experiencing supraventricular arrhythmias. Even in this setting, long-term use should be frequently reevaluated as the patient ages and develops cardiac disease. If there is a need for in-hospital initiation of drug because of significant incidence of proarrhythmia even in the absence of cardiac pathology, there would be little hope that the new antiarrhythmic drug would be accepted, either by the regulatory boards or by the medical community. In contrast, if decreased or an unaffected mortality rate is demonstrated, this would permit drug use for atrial arrhythmias in patients with moderate LV dysfunction (EF 36-50%).

IV. CONCLUSIONS

Clinical trials of potentially useful antiarrhythmic drugs are at the present time both complex and costly. For the time being, the potential market of patients with AF justifies the effort and expense to complete such studies in order to obtain regulatory board approval. This situation may improve in the future if other classes of drug action are discovered which do not cause proarrhythmia as do the Class I and III compounds. In the meantime, close attention is being paid by the pharmaceutical industry to the ongoing development of and indications for implantable defibrillators in the prevention of SD due to ventricular arrhythmias, and in the case of atrial arrhythmias, to the enormous strides being made in catheter-based ablation techniques.
V. REFERENCES

12. The Sudden Cardiac Death – Heart Failure Trial (SCD-HeFT). Late Breaking Clinical Trial Results. Annual Meeting of the American College of Cardiology, New Orleans, 2004.
Chapter 16. Epilepsies and Convulsive Disorders

Santiago Arroyo, M.D., Ph.D. ¹
David W. Chadwick, M.D. ²
Jacqueline French, M.D. ³
Richard Mattson, M.D. ⁴
Emilio Perucca, M.D., Ph.D. ⁵

¹ Medical College of Wisconsin
Froedtert Hospital
9200 West Wisconsin Avenue
Milwaukee, WI 53226
USA
Tel +1-414 805 5337
Fax +1-414 259 0469
Email sarroyo@mcw.edu

² Walton Centre for Neurology and Neurosurgery
Liverpool
UK

³ Department of Neurology
University of Pennsylvania School of Medicine
Philadelphia, PA
USA

⁴ Department of Neurology
Yale University School of Medicine
New Haven, CT
USA

⁵ Clinical Pharmacology Unit
Department of Internal Medicine and Therapeutics
University of Pavia
Pavia
ITALY

Corresponding Author:
Santiago Arroyo, M.D., Ph.D.
1. INTRODUCTORY REMARKS

With an overall prevalence in the order of 0.5 to 1%, epilepsy is a very frequent serious neurological disorder, and almost invariably requires long-term pharmacological management. Over the past 15 years, several new antiepileptic drugs (AEDs), including felbamate, fosphenytoin, gabapentin, lamotrigine, levetiracetam, oxcarbazepine, tiagabine, topiramate, vigabatrin and zonisamide, have appeared in the market with the objective of improving efficacy, tolerability, and ease of use compared with the classical AEDs (carbamazepine, valproic acid, phenytoin, phenobarbital, primidone, ethosuximide and benzodiazepines). These objectives, however, have not been completely fulfilled. Most importantly, the introduction of the new AEDs has had little impact towards the goal of achieving seizure freedom in the many patients (about one third of a typical unselected epilepsy population) who remain refractory to conventional medications.

There are many areas where clinical research is needed. The comparative efficacy and tolerability of the newer AEDs is not yet clearly defined and there is a paucity of trials comparing head-to-head these agents. Even direct comparisons between newer and older AEDs have been relatively scarce and limited to selected epilepsy syndromes (mostly patients with partial seizures with or without secondary generalization, and patients with primarily generalized tonic-clonic seizures). Moreover, most of these trials had questionable designs in terms of low sample size, suboptimal dosing regimens (often favoring the sponsor’s product), and limited duration of follow-up. Studies in generalized epilepsy syndromes and in special populations such as infants, children and the elderly have been particularly scarce.

Despite the relative abundance of different AEDs, clinicians feel that there are still unmet needs in the pharmacological treatment of epilepsy. First, over a third of the patients with epilepsy are not controlled with the current options. Second, up to a quarter of the patients exposed to a first AED will have adverse effects severe enough to require the drug’s withdrawal, and many more will suffer from chronic adverse effects limiting their quality of life. Thirdly, several epilepsy syndromes remain resistant to standard therapies. Examples include the Lennox-Gastaut syndrome and mesial temporal sclerosis. Newer drugs with improved efficacy and tolerability profiles are surely needed.

Despite the relatively “crowded” market, pharmaceutical companies should be interested in further development of newer antiepileptic compounds. Not only is epilepsy frequent and will increase further in frequency as the elderly population enlarges, but also many AEDs have found additional indications for other CNS disorders (e.g., migraine prophylaxis, neuropathic pain, anxiety, and bipolar disorder) that amplify the rewards of this line of research.

As seizures are potentially dangerous events, placebo as the only treatment has generally not been accepted in AED research. Therefore, AEDs are investigated initially as adjunctive therapy in patients with partial seizures refractory to available medications. However, due to the fact that newly diagnosed epilepsy is treated with a single AED, a monotherapy license is highly desirable and ultimately monotherapy trials also need to be conducted. These are usually done prior to licensing (conversion to monotherapy) or after a license for adjunctive therapy use has been obtained (monotherapy trials in patients with newly diagnosed partial and/or generalized seizures).

Clinical trials in epilepsy are largely geared towards registration in the US by the Food and Drug Administration (FDA) and in Europe by the European Medicines Evaluation Agency (EMEA). Unfortunately, registration trials leave many questions unanswered. For example, most studies are performed in either newly diagnosed patients or those with severe, refractory epilepsy, but few if any are performed in patients with less severe but established epilepsy. Certain common epilepsy syndromes, such as absence and juvenile myoclonic epilepsy are difficult to study in a randomized controlled way, and as a result there is very little useful data on the efficacy of the new AEDs in these populations. Also, certain populations, such
as the developmentally disabled, or those with psychiatric and medical comorbidities are often excluded from randomized trials, again leading to a paucity of data about safety concerns, if any, in these groups.

II. PHASE II STUDIES FOR REGISTRATION OF NEW ANTIPELLEPTIC DRUGS

II.1. Outline of a typical development plan

This phase usually begins with open label exploratory studies to assess titration rates, maximally tolerated dosages, pharmacokinetics and drug interactions. Results of these studies are critical for the appropriate design of pivotal studies. Determining the influence of other antiepileptic drugs on the study drug is particularly important, and should be investigated using well designed interaction studies. Many standard AEDs will either induce or inhibit the metabolism of a study drug, if it is cleared via hepatic microsomal enzymes. Failure to discover such an interaction in early phase II could lead to substantial underdosing or overdosing in the pivotal trial. Similarly, if the study drug inhibits the metabolism of a commonly used background AED such as phenytoin or carbamazepine, interpretation of efficacy outcome in a randomized trial could be severely confounded. Subsequently, the drug candidate is assessed against placebo as adjunctive treatment in patients refractory to standard therapies. As a rule, these studies are carried out in patients with partial seizures, because these patients are easier to recruit. Studies seek to enroll patients with a relatively high frequency of seizures, to permit evaluation of clinical response over a relatively short time scale.

Two types of randomized controlled designs have been used: cross-over and parallel. Due to difficulties in carrying out cross-over trials and problems with interpretation of results, a parallel-group design is preferred.

The typical pivotal trial will utilize a multicenter, double-blind placebo-controlled randomized, adjunctive therapy, parallel group design. At least 2 or 3 dose levels should be explored, preferably within the same trial. Early proof-of-concept monotherapy studies (presurgical design) may occasionally be included as part of the phase II development program (see section 3). Patients included in short-term studies should be allowed to enter long-term open-label follow-up.

II.2. Short-term studies

Adjunctive therapy trial in patients with refractory partial epilepsy

II.2.A. Objectives
To evaluate short-term efficacy and tolerability during adjunctive therapy use

II.2.B. Primary endpoints
a. Percentage change in seizure frequency during the treatment phase (double-blind phase, including or excluding the dose titration period) compared to baseline.
   b. Responder rate (percentage of patients with a greater than 50% reduction in seizures compared to baseline).

II.2.C. Secondary endpoints
a. Percentage change in seizure frequency per seizure type (simple partial, complex partial or secondarily generalized).
   b. Percentage of patients with seizure worsening (increase in seizures by 25% or more).
   c. Percentage of seizure-free patients.
   d. Distribution of responders (seizure reduction of 25 to 50%, 50 to 75% and >75%).
   e. Completer rate (measuring the combination of failed efficacy and tolerability).
f. In trials using different dosages, change in seizure frequency and responder rate per dosage group are to be analyzed. Usually, each dose is compared to placebo, but doses are not compared to each other.
g. Incidence and prevalence of adverse events.

II.2.D. Exploratory endpoints
a. Relationship between plasma drug concentration and dose, efficacy, and adverse effects. Drug withdrawal and rebound effects.

II.2.E. Study design
A multicenter, randomized, placebo-controlled, parallel-group design is generally used. Inclusion criteria usually require stable background treatment with one to three AEDs, a seizure frequency of at least three to four partial seizures per month, and no 28 day periods seizure free. A retrospective screening phase is used to determine the patient’s refractory status, followed by a prospective baseline during which baseline AEDs are held constant and the patient’s seizure diaries are kept. This phase should be of sufficient length to detect fluctuations in seizure frequency (usually 8 to 12 weeks). Shorter baseline periods are preferred, as longer ones retard patient recruitment. Patients are then randomized to placebo and to active treatment (with one or up to three different doses of the study medication). A titration phase of variable length is usually included depending on the characteristics of the drug (this may be crucial because many AEDs are poorly tolerated when started at full maintenance dosages). Titration flexibility might be allowed, i.e. dose may be individually down titrated in cases of poor tolerability with up to two pre-defined lower doses. A maintenance period of 8-16 weeks is then implemented during which treatment remains stable. Concomitant medications are kept stable throughout the trial. If the added agent is expected to modify plasma concentrations of concomitant AEDs, the dosages of the latter may need to be adjusted to maintain constant plasma concentrations. At the end of the trial, the patient is either withdrawn according to a pre-defined withdrawal schedule, or is changed to a predetermined dose of the study medication before unblinding in order to enter an open-label extension trial.

II.2.F. Planned sample
Assuming a two-group comparison, a sample size of about 50 to 90 patients per treatment group (depending on the population standard deviation) should permit detection of a difference of 25% between treatment groups in improvement from baseline with 80% power and a type 1 error (two-sided) of 5%.

II.2.G. Study population
Adults with refractory partial seizures, with or without secondary generalization, receiving one to three background AEDs

II.2.H. Specific inclusion criteria
a. Adults (ages 16 to 65) with partial seizures (simple, complex and secondary generalized tonic-clonic), defined according to the International League Against Epilepsy (ILAE) classification.
b. Patients should have non-controlled seizures despite a stable regimen with 1-3 established appropriate AEDs (the vagal nerve stimulator is sometimes considered as a drug).
c. A defined minimum number of seizures in the baseline phase (e.g., more than 6 observable partial seizures in 8 weeks and at least one seizure in any 4-week period during the baseline).
d. Women of childbearing potential must be using a medically acceptable method of birth control and have a negative serum HCG pregnancy test result at the initial screening visit. Oral contraceptives alone may not be considered adequate because of the potential effect of AEDs on their metabolism.

II.2.I. Specific exclusion criteria
a. Patients with generalized epilepsy syndromes.
b. Patients with a history of convulsive status epilepticus in the past year.
c. Patients with non-epileptic attacks (syncopes, pseudoseizures).
d. Patients with a clinically relevant medical illness or a significant psychiatric disorder.
e. Patients with progressive CNS disorders (vascular malformations, high grade tumors, etc.).
f. Drug or alcohol abuse.
g. Previous poor compliance with therapy.
h. Pregnant or breastfeeding women.
i. Need for rescue benzodiazepines more frequently than once in the baseline period.
j. Uncountable seizures as a result of seizure clustering, or inadequate supervision if the patient cannot count their own seizures.

II.2.J. Tools for assessing primary endpoints
Seizure diary

II.2.K. Specific criteria for early withdrawal and discontinuation
The investigator will discontinue a patient before completion of the study if consent is withdrawn, if the double-blind code is broken, if it is medically unacceptabe to continue treatment due to adverse events, seizure exacerbation, or other reasons, or if pregnancy occurs.

II.2.L. Data analysis method
The analysis of efficacy variables is based on the intention-to-treat (ITT) population (including all randomized patients who have received at least one dose of medication). All statistical tests are two-sided and p values ≤0.05 are considered statistically significant. Multiple statistical methods can be used for the primary and secondary endpoints (analysis of variance, analysis of covariance, logistic regression, Cochran-Mantel-Haenszel statistics, etc).

II.3. Long-term studies
It is customary for patients completing short-term studies to be allowed to enter long-term open-label follow-up studies. The primary objective of these studies is to provide data on tolerability and safety during long-term use and to obtain descriptive information concerning potential loss of therapeutic benefit. Typically, patients entering long-term extension trials may have the dosage of study drug and concomitant medications adjusted on the basis of clinical response. Evaluation will include an overall assessment of tolerability and seizure frequency. The study drug may be continued for as long as it is felt to be clinically beneficial, and retention of patients on the drug may be used as a crude measure of effectiveness. Additionally, extension trials may offer the opportunity to discontinue concomitant medication in order to obtain a preliminary assessment of the study drug under monotherapy conditions.

III. PHASE III STUDIES FOR REGISTRATION OF NEW ANTIEPILEPTIC DRUGS: ADJUNCTIVE THERAPY INDICATIONS

III.1. Outline of a typical development plan
During phase III development at least one large multicenter confirmatory adjunctive therapy, double-blind, randomized, placebo-controlled, parallel group study in refractory partial seizures is performed, usually with a dose-ranging design assessing two or three different doses of the study medication. Adolescents (over 12 years of age) as well as adults may be included in this study. Additional trials in pediatric populations may be performed. These may be conducted in children with refractory partial seizures and, if appropriate based on the expected spectrum of activity of the drug, also in children with refractory generalized epilepsies, such as Lennox-Gastaut syndrome or other syndromes. Monotherapy
trials in refractory patients (conversion to monotherapy design, or presurgical monotherapy design) may also be part of a phase III program: these trials are described in a separate section below.

A clinical development plan will include several different phase III studies, some of which may address special issues (e.g. effects on cognitive function and cognitive outcome, pharmacokinetics and drug interactions, efficacy and tolerability in special groups such as children, the elderly or cognitively impaired patients). The pivotal trial for registration purposes will be an adjunctive therapy trial in refractory partial seizures. Both the FDA and the EMEA will grant approval for the adjunctive therapy indication for partial epilepsy after two adequate and well controlled trials in patients with partial seizures. A trial in Lennox Gastaut syndrome is outlined below to provide an example of a study design for a different epilepsy syndrome.

III.2. Short-term adjunctive therapy studies

Adjunctive therapy trial in refractory partial epilepsy
The design for phase III adjunctive-therapy trials in refractory partial seizures is similar to that described above for phase II trials. Often, multiple doses are explored. Different regimens may also be explored, such as BID versus TID.

Adjunctive therapy trial in Lennox-Gastaut syndrome

III.2. A. Objectives:
To evaluate efficacy and short-term tolerability of adjunctive therapy in patients with Lennox-Gastaut syndrome

III.2.B. Primary endpoints
a. Determination of a statistically significant between-group difference with respect to either 1) reduction in the average monthly seizure rate for all seizure types combined or 2) each component of a compound variable consisting of percentage reduction in drop attacks (tonic-atomic seizures) and the parental global evaluation of seizure severity.
b. Reduction in the average monthly (28-day) seizure rate for all seizure types combined during the treatment phase (double-blind phase, including or excluding the dose titration period) compared to baseline.
c. Percentage reduction in drop attacks (tonic and atonic seizures).

III.2.C. Secondary endpoints
a. Percentages of patients considered to be treatment responders (defined as those with an equal or greater than 50% reduction from baseline for drop attacks, major seizures, and all seizures).
b. Reduction in the average monthly (28-day) rate of major seizures (drop attacks and tonic-clonic seizures).
c. Parental global evaluation of improvement relative to baseline.
d. Incidence and prevalence of adverse events.

III.2.D. Exploratory endpoints
a. Relationship of efficacy and tolerability parameters with dose and plasma drug concentrations. Quantitative EEG analysis.

III.2.E. Study design
A multicenter, randomized, adjunctive therapy, double-blind, placebo-controlled parallel-group design is used. The trial may consist of a 4-week baseline phase followed by a 12-week double-blind treatment phase (including a titration and a maintenance period). Only one dose (weight adjusted) may be explored.
Other aspects of the design are comparable to those described for phase II adjunctive therapy trials in refractory partial epilepsy.

**III.2.F. Planned sample**
As described above for phase II adjunctive therapy trials in refractory partial epilepsy. Sample size may be moderately lower than in refractory partial epilepsy due to the fact that patients with Lennox Gastaut syndrome tend to have greater seizure frequencies, with a lesser population standard deviation.

**III.2.G. Study population**
Male and female children (older than 4 years), adolescents and adults with Lennox-Gastaut syndrome

**III.2.H. Specific inclusion criteria**
- a. Adults and children, aged 4 to 65.
- b. History, EEG, and seizure patterns consistent with a diagnosis of Lennox-Gastaut syndrome. Seizure types will include drop attacks (i.e., tonic and atonic seizures) and either a history of or active atypical absence seizures. Other seizure types may include tonic-clonic, myoclonic, and partial-onset seizures. Seizures are classified according to the ILAE International Classification of Epilepsies and Epileptic Seizures.
- c. Patients are required to have refractory and frequent seizures during the month before entering the baseline phase while being maintained on a stable regimen with one or two standard AEDs.

**III.2.I. Specific exclusion criteria**
- As described for phase II adjunctive therapy trials in refractory partial epilepsy except, of course, for the epilepsy type.

**III.2.J. Tools for assessing primary endpoints**
Patient/caretaker seizure diaries, parental global evaluation scale.

**III.2.K. Specific criteria for early withdrawal and discontinuation**
As described for phase II adjunctive therapy trials in refractory partial epilepsy.

**III.2.L. Data analysis method**
As described for phase II adjunctive therapy trials in refractory partial epilepsy.

**III.3. Long-term adjunctive therapy studies**
Please refer to the outline described for phase II long-term adjunctive-therapy studies.

**IV. PHASE III STUDIES FOR REGISTRATION OF NEW ANTIEPILEPTIC DRUGS: MONOTHERAPY INDICATIONS**

**IV.1. Outline of a typical development plan**
As monotherapy is the standard treatment for most patients with epilepsy, approval of a monotherapy indication is very important for the success of the drug in the market place. Monotherapy studies also allow evaluation of a drug’s efficacy and tolerability profile, due to removal of the confounding effects of concomitant medication and associated drug-drug interactions.

Because the use of placebo as sole therapy is generally considered ethically unacceptable in epilepsy, most studies use an active control as comparator. This, however, may complicate the interpretation of the results. In fact, when administration of the investigational drug leads to a degree of seizure control which is comparable to that observed with the optimal standard treatment used as a reference (a realistic
scenario, given the remarkable effectiveness of established treatments in the newly diagnosed epilepsy population), the study may be regarded as lacking assay sensitivity, i.e. the two treatments might be equally ineffective in the specific patients’ population recruited in the study. To address these concerns, a number of study designs have been developed which are aimed at demonstrating a difference in favor of the investigational agent. Such protocols involve randomization of patients to a high dosage of the investigational agent and to a suboptimal dosage of either the same agent or an established AED. The use of a suboptimal dose (sometimes referred to as “pseudoplacebo”), however, is controversial as it conflicts with the principle of equipoise which, according to the Declaration of Helsinki, should govern all clinical trials. An additional problem with trials comparing a high versus a low dosage of the investigational agent is that the design is likely to fail to identify the optimal dose range, leading to labeling specifications which may not reflect the optimal mode of use of the drug. For conversion to monotherapy trials, these problems are compounded by the fact that dosage requirements in refractory patients may not necessarily be applicable to patients with newly diagnosed epilepsy, many of whom have milder forms of the disease. With short-term trials, an additional criticism is that the endpoints used and the duration of assessment (see below) may bear little or no relevance to the therapeutic setting, where long-term seizure remission is the major objective to ensure an acceptable quality of life. For the reasons summarized above, regulatory trials which tend to rely on randomization to fixed dosages and relatively short duration of treatment do not provide the information which is required for rational prescribing. Longer duration, flexible-dosages pragmatic trials (see section 5) are suited to address these concerns.

Two different patient populations may be included in monotherapy trials: patients with refractory seizures (usually partial seizures) and patients with newly diagnosed epilepsy. The trial designs used would be different. In refractory patients two types of design are employed: the outpatient conversion to monotherapy and the in-patient presurgical withdrawal to monotherapy. Both designs involve short-term assessment aimed at demonstrating superiority over a suboptimal comparator or placebo. In newly diagnosed epilepsy two types of trials have also been applied: the superiority design, which is usually a medium-term comparison versus a suboptimal comparator or placebo, and the non-inferiority design, which typically involves a longer duration of assessment.

EMEA guidelines for granting the monotherapy indication differ somewhat from those of the FDA. The EMEA requires that the investigational drug should have proven efficacy and safety in newly diagnosed epilepsy, with use in other monotherapy situations being regarded as supportive. Non-inferiority monotherapy trials using an established comparator at optimized dosages are considered by the EMEA as the best study design, even though supportive evidence from some kind of superiority trial (conversion to monotherapy or low-dose vs. high-dose active control) is also recommended. The FDA, on the other hand, does not accept the validity of non-inferiority trials and requires clear demonstration of superiority versus a comparator, either in refractory patients (conversion to monotherapy design) or in newly diagnosed patients. An alternative monotherapy design has been proposed. It involves using historical control data derived from the many withdrawal to monotherapy outpatient trials conducted to date. According to this approach, the investigational agent is assessed at a full dosage in a conversion to monotherapy trial (without including a suboptimal treatment arm, though some other type of control, such as an established AED at a fully effective dosage, may be included for comparative purposes) and a monotherapy license will be granted if the response rate exceeds the upper limit of the confidence interval established for historical controls. In this way, use of a suboptimal treatment and related ethical concerns would be avoided. The validity of this approach is currently being considered by regulatory agencies.

Given this background, a monotherapy development plan aimed at obtaining a worldwide license currently requires at least two separate studies: a superiority trial, conducted preferably in the U.S., and a non-inferiority trial, conducted preferably in Europe.
IV.2. Short-term monotherapy studies

IV.2.i. Outpatient conversion to monotherapy trial in refractory partial epilepsy

IV.2.i. A. Objectives
To evaluate the efficacy and safety of the investigational drug as monotherapy in patients with uncontrolled partial seizures.

IV.2.i. B. Primary endpoints
Time to exit due to fulfillment of one of the exit criteria (the aim is to show that patients allocated to the high dosage of the investigational agent are less likely to experience seizure worsening compared with those allocated to a low-dose suboptimal treatment).

IV.2.i.C. Secondary endpoints
Percentage of patients meeting one of the exit criteria in each of the two treatment groups, incidence and prevalence of adverse events.

IV.2.i.D. Exploratory endpoints
Relationship of efficacy and tolerability parameters with dose and plasma drug concentrations.

IV.2.i.E. Study design
A randomized, double-blind, active control, parallel-group design is used. The trial involves a screening phase and an 8-week baseline phase, which is followed by a treatment phase including a transition period and a monotherapy period. The transition period allows for the treatments being compared to be titrated upwards and for the baseline AEDs to be progressively reduced and eventually discontinued. Withdrawal of background AEDs can be done either before or after randomization. In the more common design, patients are randomized to treatment or control, after which background AEDs are slowly withdrawn over 8-12 weeks. In the second design, all patients are converted to monotherapy treatment with the study drug in an open-label fashion, after which they are randomized to blinded treatment with high vs low dose. The transition phase is followed by a 12- to 16-week monotherapy period that includes an enriched population, i.e. all patients who have successfully converted to monotherapy and did not fulfill exit criteria (see below) in the previous phases. Dose flexibility has sometimes been allowed during this phase. At the end of the treatment phase or if there is a premature discontinuation, the investigational drug is either tapered down and substituted by another AED, or the patient enters an open-label extension phase.

IV.2.i.F. Planned sample
A sample size of about 50 patients per treatment group is required to detect a 35% difference between trial arms in the percentage of patients meeting one exit criteria with 90% power and a type 1 error (two-sided) of 5%.

IV.2.i.G. Study population
Adolescents and adults with refractory partial seizures.

IV.2.i.H. Specific inclusion criteria
Similar to those described for phase II adjunctive therapy trials.

IV.2.i.I. Specific exclusion criteria
Similar to those described for phase II adjunctive therapy trials.

IV.2.i.J. Tools for assessing primary endpoints
Patient’s seizure diary.
IV.2.i.K. Specific criteria for early withdrawal and discontinuation
The following exit criteria are defined relative to the number of seizures during the baseline: doubling of average monthly seizure rate, doubling of the highest consecutive 2-day seizure rate, emergence of more severe or new seizure types (including generalized tonic-clonic convulsions), clinically significant prolongation of generalized tonic clonic seizures.

IV.2.i.L. Data analysis method
The analysis of efficacy variables is based on the intention-to-treat (ITT) population. Time to exit is analyzed using the log-rank test and Kaplan-Meier estimates.

IV.2.ii. In-patient presurgical conversion to monotherapy trial in refractory partial epilepsy
The presurgical design is currently less favored than the conversion to monotherapy design, partly because it has been argued that this design may primarily test efficacy against drug withdrawal seizures, which may involve pathophysiological mechanisms different from those of spontaneous seizures. Moreover, the time scale over which efficacy is assessed in presurgical designs trials is perceived as bearing limited relevance for long-term clinical use. At present, performance of this trial alone would not lead to an FDA indication for use of a drug as monotherapy. It is now more often performed as a proof of principle study, during phase IIa.

This trial requires rapid introduction of the investigational drug, and therefore is only appropriate for compounds that can be initiated rapidly.

IV.2.ii.A. Objectives
To evaluate the short-term efficacy and safety of the investigational drug as monotherapy for patients with uncontrolled partial seizures. To prove that a drug enters the brain and has an antiepileptic effect (as proof of principle).

IV.2.ii.B. Primary endpoints
Time to exit due to fulfillment of one of the exit criteria. The aim is to show that patients allocated to the high dosage of the investigational agent are less likely to experience worsening of seizures compared with those allocated to a low-dose (suboptimal) treatment.

IV.2.ii.C. Secondary endpoints
Percentage of patients completing the study, percentage of patients who meet one of the exit criteria, total number of partial seizures during the double-blind phase, total number of secondarily generalized seizures, incidence and prevalence of adverse events.

IV.2.ii.D. Exploratory endpoints
Safety of quick titration/drug-loading, speed of drug action.

IV.2.ii.E. Study design
A randomized, double-blind, parallel-group design is used, with a low-dose active control or, at times, a placebo control. The study is performed in patients with refractory partial epilepsy that are admitted to hospital for video-EEG monitoring for presurgical assessment and have their AEDs withdrawn to facilitate seizure recording. Once patients have been drug-free for 48 hours, they are randomized to a high dosage of the investigational drug and to a suboptimal (low-dose or placebo) treatment. To reduce risks, a benzodiazepine (lorazepam) may be permitted during the 48 hour medication-free period and sometimes during the first 24 hours after randomization. The medication or placebo are quickly loaded and the
double-blind evaluation lasts for 8-10 days. Dose flexibility in case of adverse events may be allowed. The double-blind phase may be followed by an open-label extension study.

**IV.2.ii.F. Planned sample**
Sample size may be calculated with respect to the ability of detecting a 30% difference between the high-dosage group and the placebo/low-dose group for the percentage of patients meeting one exit criteria. If it is assumed that 85% of the placebo-treated patients meet one of the exit criteria, given a two-sided Z-test with a significance level of 0.05 and a power of 0.85, about 50 patients per group are required.

**IV.2.ii.G. Study population**
Patients with refractory partial seizures.

**IV.2.ii.H. Specific inclusion criteria**
Patients with refractory partial epilepsy undergoing AED withdrawal within a presurgical workup. Patients need to have between 2-10 partial seizures during baseline and be receiving no AEDs when randomized.

**IV.2.ii.I. Specific exclusion criteria**
Similar to those described for phase II adjunctive therapy trials in refractory partial epilepsy.

**IV.2.ii.J. Tools for assessing primary endpoints**
Video-EEG recorded seizures.

**IV.2.ii.K. Specific criteria for early withdrawal and discontinuation**
Exit criteria are predefined to ensure patient safety. These may include: 3-4 partial seizures or secondarily generalized seizures, new appearance of generalized tonic-clonic seizures, serial/prolonged seizures, or status epilepticus.

**IV.2.ii.L. Data analysis method**
The analysis of efficacy variables is based on the intention-to-treat (ITT) population. Time to exit may be analyzed using the log-rank test and Kaplan-Meier survival curves. Additional statistical analyses may be performed using a Cox’s proportional hazards regression model. Secondary efficacy variables (percentage of patients who meet one of the exit criteria) may be analyzed using the Cochran-Mantel-Haenszel test.

**IV.3. Long-term monotherapy studies**

**IV.3.i. Superiority monotherapy trial in newly diagnosed epilepsy**

**IV.3.i.A. Objectives**
To evaluate the comparative efficacy and tolerability of the investigational drug versus a (usually suboptimal) active control under monotherapy conditions in new onset epilepsy.

**IV.3.i.B. Primary endpoint**
Time to first or second seizure (seizures occurring during the titration period may or may not be censored, depending on the characteristics of the titration phase).

**IV.3.i.C. Secondary endpoints**
Time to treatment failure (discontinuation of treatment), percentage of seizure-free patients after 6 and 12 months of treatment.
IV.3.i.D. Exploratory endpoints
Relationship of efficacy and tolerability parameters with dose and plasma drug concentrations.

IV.3.i.E. Study design
A multicenter, randomized, double-blind, parallel-group study is performed in patients with untreated epilepsy using a dose-controlled design, i.e. comparing a low dosage with a high dosage. Target dosages may be reached after an appropriate titration period, and patients experiencing intolerable adverse effects during titration may be allowed to step back by one dose level. The aim of the study is to demonstrate that time to first or second seizure is longer in the high-dosage group than in the low-dosage group. The double-blind phase may be followed by an open-label extension study.

IV.3.i.F. Planned sample
It has been suggested that these studies be powered on the basis of number of failure events (a first seizure), not number of patients. Based on the hypothesis that the hazard ratio for time to first seizure is 0.525 and constant over time, 108 events are needed for 92.5% power to detect a statistically significant difference at the 5% (two-sided) level. Enrollment is to be stopped when 108 events have been observed. In the only trial performed using this approach and only allowing the inclusion of patients with 1 or 2 seizures during a retrospective 3-month baseline, a recruitment of about 500 patients was necessary.

IV.3.i.G. Study population
Patients with new onset epilepsy or previously diagnosed but currently untreated epilepsy. The population could include patients with partial seizures (with or without secondary generalization) and primarily generalized tonic-clonic seizures.

IV.3.i.H. Specific inclusion criteria
a. Age of 12 to 65 years (wider age limits may be acceptable).
b. Recently diagnosed epilepsy with two or more unprovoked seizures. It may be possible to accept one seizure plus additional unequivocal evidence supporting the diagnosis of epilepsy (epileptiform EEG activity, brain imaging evidence). Patients should have had at least one seizure within the 3 months previous to randomization (an upper limit to number of seizures during this period may be set). In addition, patients with a previous history of epilepsy that has been in remission without medications for at least 6 months and have had one seizure in the previous 3 months may be included.
c. Patients treated with a single AED for less than 2 weeks could enter the study provided that medication is withdrawn previous to randomization.
d. Women of childbearing potential must be using a medically acceptable method of birth control and have a negative serum HCG pregnancy test result at initial screening visit.

IV.3.i.I. Specific exclusion criteria
Non-epileptic attacks (syncope, pseudoseizures), history of status epilepticus, significant medical or psychiatric illness, drug abuse, progressive central nervous system disease. Depending on the drugs being compared, patients with specific epilepsy syndromes (for example, generalized epilepsies) may need to be excluded.

IV.3.i.J. Tools for assessing primary endpoints
Seizure diary.

IV.3.i.K. Specific criteria for early withdrawal and discontinuation
First seizure or adverse event requiring discontinuation of treatment.
IV.3.i.L. Data analysis method
The analysis of efficacy variables is based on the intention-to-treat (ITT) population. The primary efficacy variable is analyzed by Kaplan-Meier survival analysis. The log-rank test may be used to assess between-group differences. A Cox proportional hazards model may also be applied.

IV.3.ii. Sequential design monotherapy trial in newly diagnosed epilepsy

IV.3.ii.A. Objectives
To evaluate the comparative efficacy and tolerability of the investigational drug versus an active control (an established AED or a lower dose of the investigational drug) under monotherapy conditions in new onset epilepsy.

IV.3.ii.B. Primary endpoints
Time to first seizure following completion of the dose titration phase.

IV.3.ii.C. Secondary endpoints
Time to treatment failure (discontinuation of treatment), percentage of seizure-free patients after 6 and 12 months of treatment. Time to second, third and fourth seizure.

IV.3.ii.D. Exploratory endpoints
Relationship of efficacy and tolerability parameters with dose and plasma drug concentrations.

IV.3.ii.E. Study design
The sequential design is a trial which allows a series of interim analyses of the emerging data so that the trial can be stopped when a predetermined difference between treatments (or lack of such a difference) has been demonstrated. These trials may require fewer patients than traditional designs of equal power, and in particular can avoid continuation when one treatment is already evidently inferior to the other. As with other trials, the design involves a multicenter, randomized, double-blind, active control, parallel-group comparison. In the only large trial where this design was used in epilepsy, patients were titrated up to a predetermined target dosage and thereafter followed up for a maximum of 162 weeks. Dose adjustments were permitted if seizure control was inadequate or adverse events were observed. The double-blind phase may be followed by an open-label extension study.

IV.3.ii.F. Planned sample
The main trial conducted to date utilized a prediction of survival curves for the reference treatment (carbamazepine, initial target dose 600 mg/day) and for the investigational drug. It was regarded as possible that the investigational drug would improve the probability of surviving without a study event for 54 weeks from 0.5 to 0.6, a difference considered clinically significant. Based on the model used, a total sample size of 450 to 700 recruited patients (depending on the true seizure rates on the two treatments) was expected to be necessary to demonstrate the target difference with a power of 0.90 and at the 5% level (two-sided). This would compare with a sample size of about 1,000 patients if a fixed (non-sequential) sample design of equal power is used.

IV.3.ii.G. Study population
Patients with new onset epilepsy or previously diagnosed but currently untreated epilepsy. The population could include patients with partial seizures (with or without secondary generalization) and with primarily generalized tonic-clonic seizures (if experimental drug is broad-spectrum).

IV.3.ii.H. Specific inclusion criteria
As described above for non-sequential design monotherapy trials in newly diagnosed epilepsy.
IV.3.ii.I. Specific exclusion criteria
As described above for non-sequential design monotherapy trials in newly diagnosed epilepsy.

IV.3.ii.J. Tools for assessing primary endpoints
Seizure diary.

IV.3.ii.K. Specific criteria for early withdrawal and discontinuation
Uncontrolled seizures at the highest dosage allowed by the protocol, or adverse event requiring discontinuation.

IV.3.ii.L. Data analysis method
Each interim analysis may comprise a comparison of the survival rates on the two treatments by means of Cox’s proportional hazards regression, adjusting for seizure type and for the number of seizures during the 12 months prior to randomization. The statistics assessing the advantage of one of the treatments is denoted by the Z score, which generalizes the better known log-rank statistics to allow for any imbalance in prognostic factors. Additionally, a measure of information, denoted by V, is calculated as the null variance (approximately equal to one quarter of the total number of events). These statistics are plotted against each other at each data review, until one of the stopping boundaries of the design is crossed.

IV.3.iii. Non-inferiority monotherapy trials

IV.3.iii.A. Objectives
To evaluate the medium to long term efficacy and tolerability of an investigational drug in patients with newly onset epilepsy in comparison with an established licensed in monotherapy AED at fully effective dosages.

IV.3.iii.B. Primary endpoints
  a. Proportion of patients seizure-free for 6 months assessed in the per-protocol (PP) population.

IV.3.iii.C. Secondary endpoints
  a. Proportion of patients seizure-free for 6 months assessed in the intention-to-treat (ITT) population.
  b. Proportion of patients seizure-free for 6 months in a subset of the per-protocol (PP) population which excludes drop-out for reasons unrelated to efficacy.
  c. Percentage of patients who remain seizure-free for 12 months.
  d. Time to exit.
  e. Percentage of completers.
  f. Time to first or second seizures.
  g. Percentage of patients seizure-free at each dose.
  h. Percentage of patients withdrawn due to adverse events.

IV.3.iii.D. Exploratory endpoints
  a. Relationship of efficacy and tolerability parameters with dose and plasma drug concentrations, cognitive function measures, quality of life measures.

IV.3.iii.E. Study design
The trial may involve a multicenter, double-blind, randomized, parallel-group design comparing the investigational drug with the best reference treatment at optimized dosages. Patients are allocated to an initial target dosage of both drugs at the lower end of the expected optimal range. If the primary endpoint (6-month seizure-freedom) is not reached due to seizure recurrence after the target dose has been attained, the patient is up-titrated to a higher pre-determined dosage. If the primary endpoint (6-month seizure-freedom) is again not reached due to seizure recurrence, the patient is up-titrated to the highest dosage.
level. Patients may be allowed to step back to an intermediate dosage if side effects are encountered during each of the titration phases. The double-blind phase may be followed by an open-label extension study.

**IV.3.iii.F. Planned sample**
Assuming a 6-month seizure-free rate of 45% with the reference comparator, a true difference of zero between treatments and a 20% rate of protocol violators, a total sample size of 580 recruited patients would be required to ensure a lower limit above −15% for the two sided 95% confidence interval for the difference in 6-month seizure-free rates, with 90% power.

**IV.3.iii.G. Study population**
Patients with new onset epilepsy or previously diagnosed but currently untreated epilepsy. The typical population could include patients with partial seizures (with or without secondary generalization) and with primarily generalized tonic-clonic seizures, if the investigational drug is broad-spectrum.

**IV.3.iii.H. Specific inclusion criteria**
As described above for non-sequential design superiority monotherapy trials in newly diagnosed epilepsy.

**IV.3.iii.I. Specific exclusion criteria**
As described above for non-sequential design superiority monotherapy trials in newly diagnosed epilepsy.

**IV.3.iii.J. Tools for assessing primary endpoints**
Seizure diary

**IV.3.iii.K. Specific criteria for early withdrawal and discontinuation**
Seizures uncontrolled at the highest dosage level, adverse events requiring discontinuation of treatment.

**IV.3.iii.L. Data analysis method**
In non-inferiority trials, analysis of the primary efficacy variable is made on the per-protocol (PP) population. Six-month seizure-free rates may be compared by a logistic regression model whose 95% confidence interval computation may include treatment and seizure types (e.g. partial vs. generalized tonic-clonic seizures without clear focal onset) as factors. Interactions between treatment group and seizure types may be excluded by applying an additional logistic regression model including treatment, seizure type and treatment by seizure type factors.

**V. OTHER STUDIES (SPECIAL INDICATIONS, PRAGMATIC TRIALS)**

**V.1. Paediatric epilepsies**
Studies of AEDs in children are conducted in three main patient populations:

a. Patients with refractory partial seizures. For this population, a multicenter, randomized, double-blind, adjunctive therapy, parallel-group, placebo-controlled trial may be performed using a design similar to that described for adults;

b. Patients with refractory generalized epilepsy. For this population, a multicenter randomized, double-blind, adjunctive therapy, parallel-group, placebo-controlled trial similar to that described in section 3 for Lennox-Gastaut syndrome may be performed. Studies in certain syndromes may require specific protocols: for example, trials in absence epilepsy may be of shorter duration and should use EEG changes (e.g., reduction/disappearance of spike-and wave activity ) as primary efficacy endpoint;

c. Patients with newly diagnosed partial or generalized epilepsy. These studies usually involve multicenter, randomized, double-blind, monotherapy, active control trials, which are initiated after evidence of efficacy has been obtained from adjunctive therapy trials. The range of designs is
similar to those described for monotherapy trials in adults. Not uncommonly, inclusion criteria for monotherapy trials in newly diagnosed patients allow simultaneous inclusion of children and adults. Studies in certain paediatric epilepsy syndromes, however, may require syndrome-specific protocols.

Both the FDA and the EMEA indicate that safety data in paediatric populations should be included in the registration dossier. The EMEA suggests that paediatric studies should be initiated as early as the development program allows and that a minimum of 100 children should be followed-up for at least one year. Moreover, it is recommended that short-term and long-term studies be designed to detect possible impact on learning, intelligence, growth, endocrine functions and puberty. Paediatric pharmacokinetic data are also required.

V.2. Epilepsies in the elderly

Pharmacokinetic and safety data in a reasonable number of elderly patients (100 or more) should be collected during phase III. This should include an evaluation of potential effects on cognitive function and sedation, as well as interactions with medications frequently used in this age group. Randomized monotherapy trials in elderly patients may be performed to obtain information supporting the use of a new AED in this segment of the population.

V.3. Acute repetitive seizures and status epilepticus

Acute repetitive seizures and status epilepticus are emergency situations where acute treatment, usually by the parenteral route, is indicated. Special trial designs are required for these indications. These involve multicenter randomized active control trials, though in some circumstances (which exclude convulsive status epilepticus) use of placebo may be justifiable. Examples of trial designs for these conditions can be found among the landmark trials listed in section 7.

V.4. Pragmatic trials

Pragmatic trials are designed to reproduce conditions which more closely reflect the use of a drug in routine clinical practice. Randomized pragmatic trials may be designed to assess the relative value of different therapeutic strategies rather than individual drugs (for example, early versus deferred treatment in those situations where the indication to treat is in doubt) or to compare two or more AEDs under conditions which allow physicians to optimize dosages and other treatment modalities according to personal clinical judgement. These studies may be of little value to regulators, but they provide useful information on which to guide rational prescribing. Most of these trials follow a design similar to the randomized non-inferiority active-control monotherapy trial outlined in section 4, but they allow greater dosing flexibility and use more relevant clinical endpoints (e.g., 12-month seizure remission rates, and retention on the allocated treatment). One example of such studies is the landmark Veterans Administration trial which compared phenobarbital, primidone, carbamazepine and phenytoin in patients with partial and/or secondarily generalized tonic-clonic seizures.

VI. EXAMPLES OF LANDMARK WELL DESIGNED TRIALS

Adjunctive therapy, partial epilepsy

**Adjunctive therapy, Lennox-Gastaut-syndrome**

**Monotherapy, short-term trials**

**Monotherapy, long-term trials**

**Acute repetitive seizures and status epilepticus**

**VII. SUGGESTED READINGS**


Chapter 17. Headache Disorders

T.J. Steiner, Ph.D. ¹

Peer Tfelt-Hansen, M.D., Ph.D. ²

Douglas C. McCrory, M.D. ³

¹ Division of Neuroscience
Imperial College London
St. Dunstan’s Road
London W6 8RP
UK

² Department of Neurology
University of Copenhagen
Glostrup Hospital
DK-2600 Glostrup
DENMARK

³ Center for Clinical Health Policy Research
Department of Medicine
Duke University Medical Center and the Center for Health Services Research in Primary Care
Durham Veterans Affairs Medical Center
Durham, NC
USA
I. INTRODUCTORY REMARKS

Headache disorders are ubiquitous. Their lifetime prevalence in populations in which it has been measured is over 90%. They are also disabling, and costly.

*Migraine* manifests in attacks lasting a few hours to three days, with a median frequency of one per month. It has a 1-year prevalence of 12-15% throughout Europe and in North and South America. In other parts of the world there is gathering evidence of similar prevalence. In Japan it is estimated to affect 8.4% of the adult population. Elsewhere in the Far East, surveys of representative samples of the general population are difficult; prevalence and impact, which have been thought to be lower, may be underestimated. In an American survey 80% of people with migraine reported some degree of disability because of it and globally, according to WHO’s *World Health Report 2001*, migraine is 19th in the list of all causes of years lived with disability. Because migraine affects people particularly during their productive years, its economic impact is high. Additionally, as well as suffering directly from its symptoms, people with migraine consistently score highly on scales of general physical and mental ill-health.

*Tension-type headache* is the most common headache disorder. Most is episodic, with occasional attacks lasting hours of what sufferers often describe as “ordinary headache”. In its chronic and much more disabling sub-type, on the other hand, it is present on more days than not. *Chronic tension-type headache* overlaps with and is sometimes indistinguishable from other forms of *chronic daily headache*, some of which are unrelentingly present throughout every day. Estimates of the prevalence of this group of conditions in Europe and USA are as high as 1 in 25 of the entire adult population. Such frequent headache is associated with long-term morbidity and disability.

*Cluster headache* has a lower prevalence (lifetime about 0.07%) and at any given time most people with the disorder are not in a cluster period. It is unique amongst the primary headache disorders in affecting men more than women (about 6:1). Typically occurring in bouts of a few weeks each year with periods of full remission between, it is characterised by frequent (daily or more often) short-lasting (15-120 minutes) but excruciating unilateral localised frontal or peri-orbital pain accompanied by marked but similarly localised autonomic symptoms. In its rarer chronic sub-type there are no periods of remission.

The financial cost of headache arises partly from direct treatment costs but, to a much greater extent, from consequential losses: work time and productivity losses are by far the largest elements. That these costs remain high throughout the world is evidence of treatment failure, a problem attributable in part to the fact that available drugs fall well short of being ideal treatments for any of these headache disorders. In another important part, it is due to the low priority generally given to headache in the queue for health-care resource allocation.

Drugs are therefore required that are not only more effective but also shown to be cost-effective in the relief of headache. Surprisingly, pharmaco-economic studies in this area are in their infancy. Whilst research is needed to derive simple agreed cost-of-illness measures that adequately capture those aspects of cost that matter to patients, time is an important casualty of headache and time losses (and their reduction by effective treatment) should be relatively easy to measure.

Amongst the primary headaches, only migraine has benefited substantially from recent pharmaceutical investment. The result has been the marketing since 1991 of seven triptans, a class of drug which has unquestionably moved forward the acute treatment of migraine whilst proving to be of some value, albeit limited, in cluster headache also. Yet triptans are far from being 100% efficacious, and they are not universally tolerated. Whilst much current research concentrates on head-to-head comparisons between triptans, invariably showing minor differences, development of new drug classes may have stalled. Meanwhile, symptomatic treatments (analgesics and anti-emetics) remain very useful in acute migraine.
therapy. They are still the mainstay of treatment of migraine in children and for many adults, including those for whom triptans are not appropriate, and they are the only option in the large parts of the world that new and relatively expensive drugs do not penetrate. They are, generally, the first-line treatments for tension-type headache.

The triptan successes displaced interest from preventative studies. For all primary headache disorders, prophylactic drugs now available are, at best, of quite limited value.

Clearly there is much unfulfilled therapeutic need in all of these areas requiring further clinical research. For pharmaceutical companies, potential markets are very large indeed although experience shows that the greater part of these is not easily penetrated on the basis of efficacy studies alone.

II. PHASE II STUDIES FOR REGISTRATION OF NEW DRUGS

II.1. Outline of a typical development plan

Whilst the requirements differ for migraine, tension-type headache and cluster headache, and in each of these for acute and preventative therapies, there are general principles applying to all.

During phase II, candidate drugs are assessed for efficacy against placebo. Randomised controlled trials will follow pharmacokinetic studies. For drugs intended for the acute treatment of migraine, pharmacokinetic studies must be conducted during the attack because absorption may be delayed by gastric stasis. For this reason oral administration, though preferred by most patients, is not ideal in acute migraine: early proof-of-concept studies may use parenteral therapy if there is doubt about rapid bioavailability by the oral route and alternative routes may anyway be required in severely nauseated or vomiting patients. Cluster headache attacks typically have duration of 15-120 minutes; rapid bioavailability is needed if treatment is to have worthwhile effect, and this is unlikely to be achieved by the oral route. Phase II must address formulation issues.

Recent experience in migraine has shown that phase II often needs to incorporate dose-finding studies, and the EMEA now require to know both the lower end of the clinically effective dose range and the optimal dose(s) (inter-individual variation may be high enough that a range of doses should be marketed). Typical studies will be double-blind, involving one or up to three or more doses. They may use parallel-groups or (multiple) cross-over designs, with regulators generally favouring the former.

Regulators have not so far distinguished, for drug-development purposes, between episodic tension-type headache and other causes of mild-to-moderate pain.

Acute-treatment trials in phase II usually allow the treatment of a single attack. Attrition rates tend to be high between attacks (one of the arguments against cross-over designs).

Prophylactic trials must allow time for dose-titration (if needed) and then time for effect to develop and be measurable, usually requiring a minimum of three months’ treatment for migraine or chronic tension-type headache. The objectives of treatment in these two disorders are not identical: migraine prophylaxis is intended to reduce the frequency of continuing attacks; in chronic tension-type headache the intent is to cause reversion to the episodic subtype. Patients with cluster headache, where use of placebo as a control may be difficult, will expect rapid and obvious efficacy achieving attack suppression or, ideally, attack cessation. Furthermore, in episodic cluster headache, spontaneous remission of the cluster period attenuates a trial’s ability to detect true treatment benefits. Any need for a period of dose-titration to balance efficacy and tolerability is a significantly complicating factor in cluster headache trial design. Long-term continuation protocols in cluster headache are not part of phase II.
Most headache disorders are appropriately treated in primary care and arguably that is where studies should be done, but this may not be true of phase II. Nevertheless, at this stage of development patients should be selected who are typical of the disorder, particularly avoiding those with complicated or refractory presentations who tend to accumulate in specialist headache centres. Exclusion criteria commonly markedly constrain recruitment, and in most cases multiple centres are needed to reach recruitment targets within reasonable timeframes.

Pharmacokinetic interactions between the test agent and other drugs likely to be taken in clinical practice should be evaluated early on. The majority of headache patients are women of child-bearing potential so oral contraceptives are important amongst these. Teratogenicity should also be assessed early: not only will teratogenic drugs be of little value in headache but also, with the notable exception of cluster headache, it is extremely difficult to recruit to headache studies if women of child-bearing potential must be excluded.

II.2. Short term studies

II.2.i. Acute treatment of migraine

II.2.i.A. Objectives

To evaluate efficacy in relieving symptoms of the acute attack.

II.2.i.B. Primary end-point

a. “Pain-free” rate: percentage of patients pain-free at 2 hours after treatment.

II.2.i.C. Secondary endpoints

a. “Headache-relief” rate: percentage of patients with a decrease in headache intensity from severe or moderate to mild or no pain within 2 hours after treatment (this was widely adopted as the primary end-point in past studies and remains important in phase II for purposes of comparison).

b. Percentage of patients pain-free at 2 hours and remaining pain-free, without use of further medication, for 48 hours after treatment.

c. Rate of relapse, defined as the return of headache of any intensity within 48 hours in patients pain-free at 2 hours after treatment.

d. Headache intensity at various time points after treatment.

e. Functional disability on a validated scale (usually a 4-point verbal rating scale where 0 = no functional impairment, 1 = “can do everything albeit with difficulty”, 2 = “cannot do some things” and 3 = “cannot do anything and/or bed-bound”) at 2 hours and other time points after treatment.

f. Effect on associated symptoms such as nausea, vomiting, photophobia and phonophobia.

g. Rate and timing of use of rescue medication.

h. Incidence and nature of adverse events.

II.2.i.D. Study design

These are invariably short-term studies. Recommended are randomised, double-blind, placebo-controlled parallel-groups studies treating one attack per patient. Three-arm trials, including placebo, are required for internal validation in active-comparator studies because of the large and highly variable placebo effect in migraine studies. In general, there is no need for stratification. Treatment may, depending on its nature, require the patient to attend a treatment facility such as the doctor’s office or be taken by the patient at home or wherever he or she may be. The former is logistically difficult. In the latter case, outcome variables are usually recorded by the patient in paper or electronic diaries, with prompts at various time points. Rescue medication should be allowed after 2 hours (for this reason, attack duration is not considered a useful secondary endpoint). The observation period after treatment of an attack should be 48 hours. Outpatients should return for final review soon after this.
II.2.i.E. Planned sample
Sample size calculations should be based on demonstrating superiority over placebo in a primary analysis of difference in pain-free rates, with an absolute difference of 20% being clinically significant and allowing for a placebo rate of up to 20%.

II.2.i.F. Study population
Adults with migraine with or without aura.

II.2.i.G. Specific inclusion criteria
a. Patients with migraine conforming to IHS diagnostic criteria 1.1 or 1.2 for at least 1 year and with at least 3 months’ well-documented retrospective history.
b. Migraine attacks occurring 1-6 times monthly.
c. Males and females.
d. Unless otherwise justified, patients should be over 18 years of age.

At the time of treatment:
a. An acute attack, usually with onset within the previous 12 hours.
b. At least 48 hours since resolution of the previous attack.
c. Headache of moderate or severe intensity.
d. So far untreated.

II.2.i.H. Specific exclusion criteria
a. Age at onset of migraine of 50 years or over.
b. Other headaches not well distinguished from migraine or occurring with such frequency as to interfere with assessments (usually taken to mean occurring on >6 days/month).
c. Other illnesses likely to interfere with assessments.
d. Use of migraine prophylactic drugs in the previous month.
e. Use of or requirement for other unacceptable concomitant therapy.
f. History of drug or alcohol overuse.

II.2.i.I. Tools for assessing endpoints
Paper or electronic diaries, with prompts at various time points.

II.2.i.J. Data analysis method
In early efficacy studies, explanatory (per protocol) analysis may be appropriate. It is unhelpful at this stage to include patients with major protocol violations. Standard statistical methods are appropriate. Adverse events are usually analysed descriptively.

II.2.ii. Acute treatment of episodic tension-type headache

II.2.ii.A. Objectives
To evaluate efficacy in relieving pain and functional impairment attributable to acute episodic tension-type headache.

II.2.ii.B. Primary end-point
a. “Pain-free” rate: percentage of patients pain-free at 2 hours after treatment.

II.2.ii.C. Secondary endpoints
a. Headache intensity (scored on either a visual analogue scale or a 4-point verbal rating scale [0 = no pain; 1, 2, 3 = mild, moderate, severe pain]) at 2 hours and other time points after treatment.
b. Headache intensity difference (the arithmetic change from baseline in headache intensity score) at 2 hours and at other time points after treatment.
c. Headache relief (on a verbal rating scale from “none” to “complete”, with two or more intermediaries which may include “meaningful relief”; negative scores may be incorporated to indicate worsening) at 2 hours and at other time points after treatment.
d. Functional disability on a validated scale (e.g., a 4-point verbal rating scale where 0 = no functional impairment, 1 = “can do everything albeit with difficulty”, 2 = “cannot do some things” and 3 = “cannot do anything and/or bed-bound”) at 2 hours and other time points after treatment.
e. Rate and timing of use of rescue medication.
f. Incidence and nature of adverse events.

II.2.ii.D. Study design
These are invariably short-term studies. Recommended are randomised, double-blind, placebo-controlled parallel-groups studies treating one attack per patient. Three-arm trials, including placebo, are required for internal validation in active-comparator studies because of the very large placebo effect reported in acute episodic tension-type headache studies. There is no need for stratification. Treatment is taken by the patient at home or wherever he or she may be. Outcome variables are usually recorded by the patient in paper or electronic diaries, with prompts at various time points. Rescue medication should be allowed after 2 hours. The observation period after treatment should be at least 24 hours. Patients should return for final review soon after this.

II.2.ii.E. Planned sample
Sample size calculations should be based on demonstrating superiority over placebo in a primary analysis of difference in pain-free rates, with an absolute difference of 20% being clinically significant and allowing for a placebo rate of up to 50%.

II.2.ii.F. Study population
Adults with episodic tension-type headache drawn (by advertising if necessary) from the general population (this is not a disorder that usually causes medical consultation; if it does, this is probably because of complicating factors or comorbidity).

II.2.ii.G. Specific inclusion criteria
a. Patients with frequent episodic tension-type headache (occurring on >1 but <15 days per month) conforming to IHS diagnostic criteria 2.2 for at least 1 year and with at least 3 months’ well-documented retrospective history.
b. Usual headache duration at least 4 hours.
c. Males and females.
d. Unless otherwise justified, patients should be over 18 years of age.

At the time of treatment:
a. An acute episode of tension-type headache, usually with onset within the previous 12 hours.
b. Headache of at least moderate intensity.
c. So far untreated.

II.2.ii.H. Specific exclusion criteria
a. Age at onset of tension-type headache of 50 years or over.
b. Chronic tension-type headache.
c. Other headaches, especially migraine and medication-overuse headache.
d. Other illnesses likely to interfere with assessments.
e. Use of prophylactic drugs in the previous month.
f. Use of or requirement for psychotropic medication or other unacceptable concomitant therapy.
g. History of drug or alcohol overuse.

II.2.ii.I. Tools for assessing endpoints
Paper or electronic diaries, with prompts at various time points.

II.2.ii.J. Data analysis method
In early efficacy studies, explanatory (per protocol) analysis may be appropriate. It is unhelpful at this stage to include patients with major protocol violations. Standard statistical methods are appropriate. Adverse events are usually analysed descriptively.

II.2.iii. Acute treatment of episodic or chronic cluster headache

II.2.iii.A. Objectives
To evaluate efficacy in aborting or suppressing the acute attack.

II.2.iii.B. Primary end-point
a. “Aborted attack” rate: percentage of patients in whom the attack is effectively stopped (headache intensity reduced to mild or no pain) within a prescribed time interval (which may be as short as 10 minutes).

II.2.iii.C. Secondary endpoints
a. Time to meaningful relief.
b. Time to “complete” relief (mild or no pain).
c. Rate of relapse, defined as the return of headache of moderate or greater intensity within 1 hour in patients reporting an aborted attack.
d. Rate and timing of use of rescue medication.
e. Incidence and nature of adverse events.

II.2.iii.D. Study design
The dramatic nature of cluster headache attacks and low placebo-response rates make open and single-blind trials informative as pilot studies. Formal comparisons with placebo must follow. These are invariably short-term randomised, double-blind, placebo-controlled studies with parallel-groups or crossover design. The latter may be preferable in this disorder since patients are uncommon but attacks occur with high frequency. In phase II, episodic and chronic cluster headache should probably be separated; if they are not, stratification is recommended in parallel-groups studies because responses to treatment may differ. Stratification for gender is also recommended for the same reason. Each patient should treat or be treated for one attack with study medication. Treatment may, depending on its nature, require the patient to be admitted to hospital, or attend a treatment facility such as the doctor’s office at a time when an attack is anticipated, or it may be taken by the patient at home or wherever he or she may be. Outcome variables are usually recorded by the doctor or patient in paper or electronic diaries, with prompts at various time points. Rescue medication should be allowed after the time prescribed for the primary end-point, but options for this are very limited. The observation period after treatment of an attack should be not less than 24 hours unless interrupted by the occurrence of the next attack. Outpatients should return for final review within 2 days.

II.2.iii.E. Planned sample
Sample size calculations should be based on demonstrating superiority over placebo in a primary analysis of difference in aborted-attack rates. Placebo-response rate is low but, because of the short attack-duration, spontaneous remission rates may be high; the two may combine to 50%. An absolute difference of 25% is clinically significant.
II.2.iii.F. Study population
Adults with episodic or chronic cluster headache.

II.2.iii.G. Specific inclusion criteria
a. Patients with episodic or chronic cluster headache conforming to IHS diagnostic criteria 3.1 or 3.2; patients with episodic cluster headache should be in at least their second cluster period.
b. Acute attacks occurring between once every 2 days and 5 times per day.
c. Attack duration of 30-180 minutes.
d. Males and females.
e. Unless otherwise justified, patients should be over 18 years of age.

At the time of treatment:
a. An acute attack, usually with onset within the previous 15 minutes (at least 15 minutes before expected spontaneous resolution).
b. At least 1 hour since resolution of the previous attack and 24 hours (or 5 half-lives if longer) since the latest previous use of study drug.
c. Headache of moderate or greater intensity.
d. So far untreated.

II.2.iii.H. Specific exclusion criteria
a. Other headaches not well distinguished from cluster headache.
b. Other illnesses likely to interfere with assessments.
c. Concurrent use of prophylactic drugs for cluster headache.
d. Use of or requirement for other unacceptable concomitant therapy.
e. History of drug or alcohol overuse.

II.2.iii.I. Tools for assessing endpoints
Paper or electronic diaries, with prompts at various time points.

II.2.iii.J. Data analysis method
In early efficacy studies, explanatory (per protocol) analysis may be appropriate. It is unhelpful at this stage to include patients with major protocol violations. Subgroup analyses (for episodic and chronic subtypes and for gender differences) are recommended and should be specified a priori. Standard statistical methods are appropriate. Adverse events are usually analysed descriptively.

II.3. Long term studies

II.3.i. Migraine prophylaxis

II.3.i.A. Objectives
To evaluate efficacy in migraine-attack prevention.

II.3.i.B. Primary end-points
a. Frequency of attacks per specified unit time (usually 4 weeks) measured during treatment after a specified period (usually 8 weeks).
b. Response rate: percentage of patients with frequency reduction of 50% or more after a specified treatment period.

The number of attacks should be recorded irrespective of their duration, and the following rules distinguish an attack of long duration from two attacks and between attacks and relapses:
a. A migraine attack which is interrupted by sleep, or which temporarily remits spontaneously and then recurs within 48 hours after its onset, should be recorded as one attack and not two.
b. An attack treated successfully with medication but with relapse within 48 hours counts as one attack.

II.3.i.C. Secondary endpoints
a. Frequency of attacks over the entire treatment period.
b. “Migraine days” (defined as any day on which symptoms of migraine are present) per 4 weeks.
c. Intensity of migraine headache averaged over attacks within a specified evaluation period.
d. Drug consumption for symptomatic or acute treatment totalled over an evaluation period.
e. Incidence and nature of adverse events.

II.3.i.D. Study design
These are invariably medium-term studies (at least 4 months) conducted in outpatients. Recommended are randomised, double-blind, placebo-controlled parallel-groups studies. Three-arm trials, including placebo, are required for internal validation with active comparators because of the large and highly variable placebo effect in prophylactic migraine studies. Randomisation should occur after a run-in (baseline) period of at least one month, when stratification for baseline attack rate (e.g., ≥3 or <3 per 4 weeks) is recommended as the prophylactic effect may depend on this variable. Treatment periods should be at least 3 months. Patients should take their usual acute therapy as required, provided that it can be safely administered with the study drug. Attacks (and, if required, their features), acute medication use and adverse events should be recorded as they occur by patients in paper or electronic diaries. Reviews of patients and their diaries should be undertaken every 4 weeks.

Compliance with medication regimens, and concordance, are known to be highly suspect in migraine prophylaxis. Counts of returned medication are unreliable for detecting poor concordance, which may render an efficacious drug useless. In phase II it is especially important to ascertain that the drug has been taken as prescribed. Consideration should be given to using electronic event monitors.

II.3.i.E. Planned sample
Sample size calculations should be based on demonstrating superiority over placebo in a primary analysis of a) difference in attack frequencies, with a relative difference of 50% or an absolute difference of 1 attack/month being clinically significant and allowing for a reduction on placebo of up to 30% or 1 attack/month; or b) difference in responder rates, with an absolute difference of 20% being clinically significant and allowing for a placebo rate of up to 30%.

II.3.i.F. Study population
Adults with frequent attacks of migraine with or without aura.

II.3.i.G. Specific inclusion criteria
a. Patients with migraine conforming to IHS diagnostic criteria 1.1 or 1.2 for at least 1 year and with at least 3 months’ well-documented retrospective history.
b. Migraine attacks occurring 2-6 times monthly.
c. Males and females.
d. Unless otherwise justified, patients should be over 18 years of age.

II.3.i.H. Specific exclusion criteria
a. Age at onset of migraine of 50 years or over.
b. Other headaches not well distinguished from migraine or occurring with such frequency as to interfere with assessments (usually taken to mean occurring on >6 days/month).
c. Other illnesses likely to interfere with assessments.
d. Use of other migraine prophylactic drugs in the previous month.
e. Use of or requirement for other unacceptable concomitant therapy.
f. Risk of pregnancy.
g. History of drug or alcohol overuse.

II.3.i.I. Tools for assessing endpoints
Paper or electronic diaries.

II.3.i.J. Specific criteria for early withdrawal and discontinuation
a. Withdrawal of consent (which may be related to lack of efficacy or adverse events).
b. Medical concerns related to evident lack of efficacy or adverse events.
c. Other intercurrent illness.
d. Pregnancy.
e. Breach of double-blinding.

II.3.i.K. Data analysis method
Even in early efficacy studies of prophylaxis, explanatory (per protocol) analysis may be misleading. Whilst it is unhelpful at this stage to include patients with random major protocol violations, drop-outs may be treatment-related. Analysis should therefore be based on the intention-to-treat (ITT) population. Since time to onset of effect is of interest, analysis of efficacy should be for the entire treatment period as well as for the period specified by the primary endpoint (e.g., the final 4 weeks of treatment). Standard statistical methods are appropriate. Adverse events are usually analysed descriptively.

II.3.ii. Chronic tension-type headache prophylaxis

II.3.ii.A. Objectives
To evaluate efficacy against chronic tension-type headache.

II.3.ii.B. Primary endpoints
a. Number of days with headache per specified unit time (usually 4 weeks) measured during treatment after a specified period (at least 8 weeks).
b. Response rate: percentage of patients with reduction in headache days per unit time of 50% or more (implying reversion from chronic to episodic tension-type headache) after a specified treatment period.

II.3.ii.C. Secondary endpoints
a. Number of days with headache over the entire treatment period.
b. Intensity of headache on a visual analogue scale or 4-point verbal rating scale [0 = no pain; 1, 2, 3 = mild, moderate, severe pain]) averaged over attacks within a specified evaluation period.
c. Duration of headache each day.
d. Drug consumption for symptomatic or acute treatment totalled over an evaluation period.
e. Incidence and nature of adverse events.

II.3.ii.D. Study design
These are invariably medium-term studies (at least 4 months) conducted in outpatients. Recommended are randomised, double-blind, placebo-controlled parallel-groups studies. There are no licensed active comparators. Randomisation should occur after a run-in (baseline) period of at least one month during which the number of days with headache and acute or symptomatic medication consumption are recorded. Stratification is unnecessary. Treatment periods should be at least 3 months. Days with headache, intensity and duration of headache, medication use and adverse events should be recorded as
they occur by patients in paper or electronic diaries. Reviews of patients and their diaries should be undertaken every 4 weeks.

Acute medication is inappropriate treatment for this disorder and should not be encouraged (regular use of acute or symptomatic medication on >2 days per week will put the diagnosis in question as this approaches the threshold for medication-overuse headache).

Compliance with preventative medication has not been evaluated in chronic tension-type headache. It may be better than in migraine because symptoms are present daily or on most days rather than intermittently. Nevertheless, in phase II it is important to ascertain that the drug has been taken as prescribed. Consideration should be given to using electronic event monitors.

II.3.ii.E. Planned sample
Sample size calculations should be based on demonstrating superiority over placebo in a primary analysis of difference in days with headache. A change from baseline of $\geq 50\%$ represents a clinically significant benefit of treatment, but the response to placebo has not been well documented (a reduction of up to 30$\%$ should be anticipated). Alternatively the primary analysis may be of difference in response rates. An absolute difference of 20$\%$ would be clinically significant. Again the response rate to placebo has not been well documented but up to 30$\%$ should be anticipated.

II.3.ii.F. Study population
Adults with chronic tension-type headache drawn from secondary or primary care or from the general population.

II.3.ii.G. Specific inclusion criteria
a. Patients with chronic tension-type headache conforming to IHS diagnostic criteria 2.3 for at least 3 months and with at least 3 months’ well-documented retrospective history.
b. Males and females.
c. Unless otherwise justified, patients should be over 18 years of age.

II.3.ii.H. Specific exclusion criteria
a. Age at onset of chronic tension-type headache of 50 years or over.
b. Other headaches, especially migraine, not well distinguished from tension-type headache or occurring with such frequency as to interfere with assessments.
c. Other illnesses, particularly depression, likely to interfere with assessments.
d. Use of other prophylactic drugs in the previous month.
e. Use of acute or symptomatic medication for headache on an average of >2 days per week over the previous 2 months.
f. Other history of drug or alcohol overuse.
g. Use of or requirement for psychotropic medication or other unacceptable concomitant therapy.
h. Risk of pregnancy.

II.3.ii.I. Tools for assessing endpoints
Paper or electronic diaries.

II.3.ii.J. Specific criteria for early withdrawal and discontinuation
a. Withdrawal of consent (which may be related to lack of efficacy or adverse events).
b. Medical concerns related to evident lack of efficacy or adverse events.
c. Other intercurrent illness.
d. Pregnancy.
e. Breach of double-blinding.
II.3.ii.K. Data analysis method
Even in early efficacy studies of prophylaxis, explanatory (per protocol) analysis may be misleading. Whilst it is unhelpful at this stage to include patients with random major protocol violations, drop-outs may be treatment-related. Analysis should therefore be based on the intention-to-treat (ITT) population. Since time to onset of effect is of interest, analysis of efficacy should be for the entire treatment period as well as for the period specified by the primary endpoint (e.g., the final 4 weeks of treatment). Standard statistical methods are appropriate. Adverse events are usually analysed descriptively.

II.3.iii. Prophylaxis of episodic cluster headache

II.3.iii.A. Objectives
To evaluate efficacy in terminating a cluster period or in reducing frequency, intensity and/or duration of continuing cluster headache attacks.

II.3.iii.B. Primary end-points
a. Frequency of attacks per specified unit time (usually 1 week) measured during treatment after a specified period (to allow treatment effect to develop) following dosage-stabilisation.

b. Remission rate: percentage of patients whose attacks have ceased after a specified treatment period.

The number of attacks should be recorded irrespective of their intensity or duration. An attack treated successfully with acute medication but with relapse within 1 hour counts as one attack.

II.3.iii.C. Secondary endpoints
a. Frequency of attacks over the entire treatment period.

b. Time to remission.

c. Intensity of cluster headaches averaged over a specified evaluation period.

d. Duration of cluster headaches summed or averaged over a specified evaluation period.

e. Drug consumption for symptomatic or acute treatment totalled over an evaluation period.

f. Incidence and nature of adverse events.

II.3.iii.D. Study design
In phase II these are short-term randomised, double-blind, placebo-controlled, parallel-groups studies conducted in outpatients. No run-in (baseline) period is needed. Stratification is recommended for time since onset of the cluster period (e.g., ≥2 or <2 weeks, but see below) and gender as each may influence the prophylactic effect or spontaneous remission rate. Treatment periods may be defined by the times prescribed for the primary end-point but should be at least 2 weeks; although they may need to incorporate dose-titration, they should not be substantially longer than this since treatments include placebo. Patients should take their usual acute therapy whenever cluster headache is of at least moderate intensity provided that it can be safely administered with the study drug. Attacks and their intensity and duration (and, if required, their associated features), acute medication use and adverse events should be recorded as they occur by patients in paper or electronic diaries. Reviews of patients and their diaries should be undertaken every week.

Compliance should be monitored. Because of the symptom frequency and severity it may be better in cluster headache than in other headache disorders but consideration should be given to using electronic event monitors.

II.3.iii.E. Planned sample
Sample size calculations should be based on demonstrating superiority over placebo in a primary analysis of a) difference in attack frequencies, with a relative difference of 50% being clinically significant and
allowing for a reduction on placebo of up to 20%; or b) difference in remission rates, with an absolute difference of 20% being clinically significant and allowing for a placebo plus spontaneous resolution rate of up to 20%.

II.3.iii.F. Study population
Adults with episodic cluster headache.

II.3.iii.G. Specific inclusion criteria
a. Patients with episodic cluster headache conforming to IHS diagnostic criteria 3.1, and in at least their second cluster period.
b. Expected duration of cluster period, from start of study medication, greater than the treatment period specified by the primary end-point (to limit the spontaneous-resolution rate in phase II, there may be advantage in restricting recruitment to patients within 2 weeks of onset of the cluster period).
c. Acute attacks occurring between once every 2 days and 5 times per day.
d. Males and females.
e. Unless otherwise justified, patients should be over 18 years of age.

II.3.iii.H. Specific exclusion criteria
a. Other headaches not well distinguished from cluster headache.
b. Other illnesses likely to interfere with assessments.
c. Other cluster headache prophylactic therapy in the previous week.
d. Use of or requirement for other unacceptable concomitant therapy.
e. Risk of pregnancy.
f. History of drug or alcohol overuse.

II.3.iii.I. Tools for assessing endpoints
Paper or electronic diaries.

II.3.iii.J. Specific criteria for early withdrawal and discontinuation
a. Withdrawal of consent (which may be related to lack of efficacy or adverse events).
b. Medical concerns related to evident lack of efficacy or adverse events.
c. Other intercurrent illness.
d. Pregnancy.
e. Breach of double-blinding.

II.3.iii.K. Data analysis method
Drop-outs may be treatment-related so, even in early efficacy studies of prophylaxis, analysis should be based on the intention-to-treat (ITT) population. Explanatory (per protocol) analysis may be worthwhile and hypothesis-generating as a secondary analysis. Subgroup analysis for gender differences is recommended and should be specified a priori. Since time to onset of effect is of interest, analysis of efficacy should be for the entire treatment period as well as for the period specified by the primary endpoint. Standard statistical methods are appropriate. Adverse events are usually analysed descriptively.

II.3.iv. Prophylaxis of chronic cluster headache
The objective in chronic cluster headache prophylaxis is long-term suppression of attacks. No methodology has yet been developed. In view of the difficulties with use of placebo as a comparator it is likely to be similar to that used in epilepsy, with add-on therapy studies preceding monotherapy trials.
III. PHASE III STUDIES FOR REGISTRATION OF NEW DRUGS

III.1. Outline of a typical development plan

During phase III, promising drugs are assessed for effectiveness against placebo in at least two pivotal large multicentre studies. They will be randomised double-blind trials incorporating one or more doses of study drug according to the findings of phase II. They may use parallel-groups or (multiple) cross-over designs, with regulators strongly favouring the former. For acute and prophylactic migraine therapy, and acute treatment of episodic tension-type headache and cluster headache, one or both of these, or additional studies in this phase, should include active comparators.

Acute treatments for cluster headache, which are unlikely to be oral, must be formulated so that patients can self-medicate wherever they may be at time of onset.

Pivotal acute-treatment trials should address not only the treatment of single attacks but also consistency of the therapeutic response across multiple attacks. Additionally in migraine and episodic tension-type headache, long-term continuation protocols are desirable to demonstrate repeatability of effect over time (lack of tachyphylaxis). Such studies contribute helpfully to safety evaluation.

In migraine particularly, the headache is not a stable pain but develops gradually, or sometimes rapidly, to a peak with subsequent spontaneous resolution. This poses challenges regarding timing of intake of medication. Trials in phase III may explore the relationship between timing of acute treatment and effect. Such a study in migraine with aura may incorporate medicating during the aura phase.

Specific trials in migraine and episodic tension-type headache may look at re-medicating, within the same attack, with a second dose of study drug when the first has been inadequately efficacious.

Prophylactic trials require a minimum of three months’ treatment for migraine or chronic tension-type headache, but are better designed to reflect treatment periods likely in routine management, which are typically longer (4-6 months or more). Furthermore, continued observation beyond the treatment period, for continuing efficacy or possibly rebound exacerbation, is essential.

Cluster headache to some extent has the status of orphan disease. Long-term prophylaxis may be inappropriate in the episodic subtype, when attacks recur over periods of only a few weeks. Some currently used drugs achieve remission quite quickly, even within a few days, and prolonged treatment may not be necessary. This, and the fact that no treatments currently used for the prevention of cluster headache are licensed for this indication, make active-comparator studies difficult whilst placebo cannot be used long-term. The regulatory requirements for phase III have not been clarified.

In all primary headache disorders, safety of treatment is a major concern since the disorders themselves are self-limiting. On the other hand they are widespread and drugs that treat them, once marketed, are likely to be taken by large numbers of people and not always in strict accordance with instructions. Regulators will look carefully at safety, and may require special studies in diseased populations.

Headache sufferers attending specialist clinics may not be representative of the larger number seen by primary-care physicians. Neither group is likely to match those in the general population who do not seek medical advice. Phase III trials need to recruit widely from the population who will use the agent when marketed, with as few restrictions as possible. Nonetheless, special protocols will be required for children (under the age of 18), whose needs may be different, and may be required for the elderly (over the age of 65), who are less subject to primary headaches and more at risk of symptomatic headache as well as other illness.
III.2. Short term studies

III.2.i. Acute treatment of migraine

III.2.i.A. Objectives
To confirm effectiveness and evaluate comparative efficacy and tolerability in relieving symptoms of the acute attack.

III.2.i.B. Primary end-point
a. “Pain-free” rate: percentage of patients pain-free at 2 hours after treatment.

III.2.i.C. Secondary endpoints
a. Percentage of patients pain-free at 2 hours and remaining pain-free, without use of further medication, for 48 hours after treatment.
b. Rate of relapse, defined as the return of headache of any intensity within 48 hours in patients pain-free at 2 hours after treatment.
c. Headache intensity at various time points after treatment.
d. “Headache-relief” rate: percentage of patients with a decrease in headache intensity from severe or moderate to mild or no pain within 2 hours after treatment.
e. Time to “meaningful” pain relief (usually defined subjectively).
f. Time to “onset of action” (defined as first noticeable pain relief).
g. Functional disability on a validated scale (usually a 4-point verbal rating scale where 0 = no functional impairment, 1 = “can do everything albeit with difficulty”, 2 = “cannot do some things” and 3 = “cannot do anything and/or bed-bound”) at 2 hours and other time points after treatment.
h. Effect on associated symptoms such as nausea, vomiting, photophobia and phonophobia.
i. Rate and timing of use of rescue medication.
j. Global evaluation of study medication.
k. Pharmacoeconomic measures.
l. Incidence and nature of adverse events.

III.2.i.D. Study design
Short-term studies should be randomised, double-blind, placebo-controlled parallel-groups outpatient trials treating one attack per patient. Three-arm trials, including placebo, are required for internal validation in active-comparator studies because of the large and highly variable placebo effect in migraine studies. In general, there is no need for stratification but phase III studies of acute treatment may opt to include patients with or without specific prophylactic medication(s), in which case stratification is based on this variable. Outcome variables are usually recorded by the patient in paper or electronic diaries, with prompts at various time points. Rescue medication should be allowed after 2 hours. This may be replaced, in patients who first received active drug, by a second dose of study drug or placebo in re-medication trials, with rescue medication deferred to 4 hours. To maintain the double-blind without subjecting patients to placebo only for 4 hours, those who first received placebo are given active drug as the second dose. The observation period after treatment of an attack should be 48 hours. Patients should return for final review soon after this.

III.2.i.E. Planned sample
Sample size calculations should be based on demonstrating superiority over placebo in a primary analysis of difference in pain-free rates, with an absolute difference of 20% being clinically significant and allowing for a placebo rate of up to 20%.

In practice, much greater numbers are required to demonstrate safety. Regulators require data from a large and representative group of patients. Trials including >1,000 patients are not unusual.
III.2.i.F. Study population
   a. Adults with migraine with or without aura.
   b. Adolescents and/or children, if they are to be included in the labelling.

III.2.i.G. Specific inclusion criteria
   a. Patients with migraine conforming to IHS diagnostic criteria 1.1 or 1.2 for at least 1 year and with at least 3 months’ well-documented retrospective history.
   b. Migraine attacks occurring 1-6 times monthly.
   c. Males and females.

At the time of treatment:
   a. An acute attack, usually with onset within the previous 12 hours.
   b. At least 48 hours since resolution of the previous attack.
   c. Headache of moderate or severe intensity.
   d. So far untreated.

III.2.i.H. Specific exclusion criteria
   a. Age at onset of migraine of 50 years or over.
   b. Other headaches not well distinguished from migraine or occurring with such frequency as to interfere with assessments (usually taken to mean occurring on >6 days/month).
   c. Other illnesses likely to interfere with assessments.
   d. Use of migraine prophylactic drugs in the previous month or, if the protocol allows inclusion of patients on prophylaxis, any change in nature or dose of prophylactic medication in the previous 3 months.
   e. Use of or requirement for other unacceptable concomitant therapy.
   f. History of drug or alcohol overuse.

III.2.i.I. Tools for assessing endpoints
   Paper or electronic diaries, with prompts at various time points.

III.2.i.J. Data analysis method
   Analysis should be based on the intention-to-treat (ITT) population, although this may be defined to exclude those known not to have taken treatment. Standard statistical methods are appropriate. Adverse events are usually analysed descriptively.

   Cost-effectiveness (or cost-utility) analysis is highly desirable, but the methodology is not yet well developed.

III.2.ii. Acute treatment of episodic tension-type headache

III.2.ii.A. Objectives
   To confirm effectiveness, and evaluate comparative efficacy and tolerability, in relieving pain and functional impairment attributable to acute episodic tension-type headache.

III.2.ii.B. Primary end-point
   “Pain-free” rate: percentage of patients pain-free at 2 hours after treatment.

III.2.ii.C. Secondary endpoints
   a. Headache intensity (scored on either a visual analogue scale or a 4-point verbal rating scale [0 = no pain; 1, 2, 3 = mild, moderate, severe pain]) at 2 hours and other time points after treatment.
b. Headache intensity difference (the arithmetic change in headache intensity score) at 2 hours and at other time points after treatment.

c. Headache relief (on a verbal rating scale from “none” to “complete”, with two or more intermediaries which may include “meaningful relief”; negative scores may be incorporated to indicate worsening) at 2 hours and at other time points after treatment.

d. Functional disability on a validated scale (e.g., a 4-point verbal rating scale where 0 = no functional impairment, 1 = “can do everything albeit with difficulty”, 2 = “cannot do some things” and 3 = “cannot do anything and/or bed-bound”) at 2 hours and other time points after treatment.

e. Rate and timing of use of rescue medication.

f. Global evaluation of study medication.

g. Incidence and nature of adverse events.

III.2.ii.D. Study design
Short-term studies should be randomised, double-blind, placebo-controlled parallel-groups outpatient trials treating one attack per patient. Three-arm trials, including placebo, are required for internal validation in active-comparator studies because of the very large placebo effect reported in acute episodic tension-type headache studies. There is no need for stratification. Outcome variables are usually recorded by the patient in paper or electronic diaries, with prompts at various time points. Rescue medication should be allowed after 2 hours. This may be replaced, in patients who first received active drug, by a second dose of study drug or placebo in re-medication trials, with rescue medication deferred to 4 hours. To maintain the double-blind without subjecting patients to placebo only for 4 hours, those who first received placebo are given active drug as the second dose. The observation period after treatment should be at least 24 hours. Patients should return for final review soon after this.

III.2.ii.E. Planned sample
Sample size calculations should be based on demonstrating superiority over placebo in a primary analysis of difference in pain-free rates, with an absolute difference of 20% being clinically significant and allowing for a placebo rate of up to 50%.

If the study drug is not already licensed for another indication, much greater numbers may be required to demonstrate safety.

III.2.ii.F. Study population
a. Adults with episodic tension-type headache drawn from the general population (by advertising if necessary).

b. Adolescents and/or children, if they are to be included in the labelling.

III.2.ii.G. Specific inclusion criteria
a. Patients with frequent episodic tension-type headache (occurring on >1 but <15 days per month) conforming to IHS diagnostic criteria 2.2 for at least 1 year and with at least 3 months’ well-documented retrospective history.

b. Usual headache duration at least 4 hours.

c. Males and females.

At the time of treatment:

a. An acute episode of tension-type headache, usually with onset within the previous 12 hours.

b. Headache of at least moderate intensity.

c. So far untreated.

III.2.ii.H. Specific exclusion criteria
a. Age at onset of tension-type headache of 50 years or over.
b. Chronic tension-type headache.
c. Migraine if not well distinguished from tension-type headache or occurring in the previous year more frequently than once per month.
d. Medication-overuse headache.
e. Other headaches not well distinguished from tension-type headache or occurring with such frequency as to interfere with assessments.
f. Other illnesses likely to interfere with assessments.
g. Use of prophylactic drugs in the previous month.
h. Use of or requirement for psychotropic medication or other unacceptable concomitant therapy.
i. History of drug or alcohol overuse.

III.2.ii.1. Tools for assessing endpoints
Paper or electronic diaries, with prompts at various time points.

III.2.ii.1. Data analysis method
Analysis should be based on the intention-to-treat (ITT) population, although this may be defined to exclude those known not to have taken treatment. Standard statistical methods are appropriate. Adverse events are usually analysed descriptively.

Longer-term studies should address consistency of the therapeutic response across attacks. These may be double-blind cross-over studies observing treatment of several attacks per patient with the same drug and dose plus, randomly, one or more (e.g., one attack out of five) with placebo. Additionally, long-term continuation protocols are desirable to demonstrate repeatability of effect over time (lack of tachyphylaxis). Such studies contribute helpfully to safety evaluation. They need be neither placebo-controlled nor blinded.

III.2.iii. Acute treatment of episodic or chronic cluster headache

III.2.iii.A. Objectives
To evaluate efficacy and comparative effectiveness and tolerability in aborting or suppressing the acute attack.

III.2.iii.B. Primary end-points
a. “Aborted attack” rate: percentage of patients in whom the attack is effectively stopped (headache intensity reduced to mild or no pain) within a prescribed time interval (which may be as short as 10 minutes).
b. Time to meaningful relief.
c. Time to “complete” relief (mild or no pain).

III.2.iii.C. Secondary endpoints
a. Rate of relapse, defined as the return of headache of moderate or greater intensity within 1 hour in patients reporting an aborted attack.
b. Headache intensity (on a 5-point verbal rating scale: 0 = no pain, 1, 2, 3, 4 = mild, moderate, severe, excruciating pain) at 5, 10 and 15 minutes after treatment and every 15 minutes thereafter for up to 3 hours (whilst these repeated assessments are recommended, marked agitation is a feature of acute cluster headache which, combined with severe pain, may make them impractical).
c. Effect on associated autonomic symptoms.
d. Functional impairment on a validated scale.
e. Rate and timing of use of rescue medication.
f. Global evaluation of study medication.
g. Patient’s preference (in cross-over studies).
h. Incidence and nature of adverse events.
III.2.iii.D. Study design
Treatments coming into phase III may be oral but are more likely to be parenteral. There are study-design implications for parenteral therapy, particularly for active-comparator studies and especially because the active comparator may itself be administered parenterally.

Short-term studies should be randomised, double-blind, placebo-controlled parallel-groups or cross-over trials in outpatients treating 1-4 attacks each. The cross-over design has advantages, and may be accepted by regulators, since patients are uncommon whilst attacks occur with high and predictable frequency. Three-arm trials, including placebo, are required for internal validation in active-comparator studies; although placebo effect is relatively slight in cluster headache, trials are easily confounded by high rates of spontaneous resolution of attacks which are short-lasting. Active-comparator studies are likely to require a double-dummy design if one or other treatment, or both, is administered parenterally. At least one large multicentre trial should formally confirm efficacy.

If a parallel-groups trial includes both episodic and chronic cluster headache, stratification is recommended because responses to treatment may differ. Stratification for gender is also recommended for the same reason. Outcome variables are usually recorded by the patient in paper or electronic diaries, with prompts at various time points. Rescue medication should be allowed after the time prescribed for the primary end-point, but options for this are very limited. The observation period after treatment of an attack should be not less than 24 hours unless interrupted by the occurrence of the next attack. Depending on the experience from phase II, reviews may shortly follow each treatment or (in multiple-attack studies) only the last treatment.

Short-term studies in cluster headache can address consistency of the therapeutic response across attacks. These may be double-blind cross-over studies, observing treatment of several attacks per patient with the same drug and dose plus, randomly, one or more (e.g., one attack out of five) with placebo.

III.2.iii.E. Planned sample
Sample size calculations should be based on demonstrating superiority over placebo in a primary analysis of difference in aborted-attack rates. Placebo-response rate is low but, because of the short attack-duration, spontaneous remission rates may be high; the two may combine to 50%. An absolute difference of 25% is clinically significant.

III.2.iii.F. Study population
a. Adults with episodic or chronic cluster headache.
   b. Adolescents and/or children, if they are to be included in the labelling.

III.2.iii.G. Specific inclusion criteria
a. Patients with episodic or chronic cluster headache conforming to IHS diagnostic criteria 3.1 or 3.2; patients with episodic cluster headache should be in at least their second cluster period.
b. Acute attacks occurring between once every 2 days and 5 times per day.
c. Attack duration of 30-180 minutes.
d. Males and females.

At the time of treatment:
a. An acute attack, usually with onset within the previous 15 minutes (at least 15 minutes before expected spontaneous resolution).
b. At least 1 hour since resolution of the previous attack and 24 hours (or 5 half-lives if longer) since the latest previous use of study drug.
c. Headache of moderate or greater intensity.
d. So far untreated.
III.2.iii.H. Specific exclusion criteria
   a. Other headaches not well distinguished from cluster headache.
   b. Other illnesses likely to interfere with assessments.
   c. Use of or requirement for unacceptable concomitant therapy.
   d. History of drug or alcohol overuse.

III.2.iii.I. Tools for assessing endpoints
Paper or electronic diaries, with prompts at various time points.

III.2.iii.J. Data analysis method
Analysis should be based on the intention-to-treat (ITT) population, although this may be defined to exclude those known not to have taken treatment. Subgroup analyses (for episodic and chronic subtypes and for gender differences) are recommended and should be specified a priori. Standard statistical methods are appropriate. “Time to” end-points require survival-analysis methods. Adverse events are usually analysed descriptively.

The frequency of medication in cluster headache may lead to high treatment costs; given that, untreated, this disorder is very disabling and ruins quality of life, cost-effectiveness (or cost-utility) analysis is appropriate but the methodology is not yet well developed.

Longer-term studies are desirable, extending throughout a cluster episode or in patients with chronic cluster headache, to demonstrate repeatability of effect over time (lack of tachyphylaxis) and for safety evaluation. These need be neither placebo-controlled nor blinded.

III.3. Long term studies

Longer-term studies should address consistency of the therapeutic response across attacks. These may be double-blind cross-over studies observing treatment of several attacks per patient with the same drug and dose plus, randomly, one or more (e.g., one attack out of five) with placebo. Additionally, long-term continuation protocols are desirable to demonstrate repeatability of effect over time (lack of tachyphylaxis). At least one study of 12 months’ duration is needed for safety evaluation. These need be neither placebo-controlled nor blinded.

One or more of these protocols should accommodate the double-blind investigation, using similar end-points, of re-medication to treat relapse following initial successful treatment with the study drug.

III.3.i. Migraine prophylaxis

III.3.i.A. Objectives
To confirm effectiveness and evaluate comparative efficacy and tolerability in migraine prevention.

III.3.i.B. Primary end-points
   a. Frequency of attacks per specified unit time (usually 4 weeks) measured during treatment after a specified period (usually 8 weeks).
   b. Response rate: percentage of patients with frequency reduction of 50% or more after a specified treatment period.

The number of attacks should be recorded irrespective of their duration, and the following rules distinguish an attack of long duration from two attacks and between attacks and relapses:
a. A migraine attack which is interrupted by sleep, or which temporarily remits spontaneously and then recurs within 48 hours after its onset, should be recorded as one attack and not two.

b. An attack treated successfully with medication but with relapse within 48 hours counts as one attack.

III.3.i.C. Secondary endpoints

a. Frequency of attacks over the entire treatment period.
b. Frequency of attacks following discontinuation of treatment.
c. “Migraine days” (defined as any day on which symptoms of migraine are present) per 4 weeks.
d. Intensity of migraine headache averaged over attacks within a specified evaluation period.
e. Speed of effect (e.g., response rates in first, second and third months of treatment).
f. Drug consumption for symptomatic or acute treatment totalled over an evaluation period.
g. Headache indices multiplying frequency, intensity and/or duration are not recommended: arbitrary weighting in the numerical scores, which may be faulty, is increased by multiplication; and indices cannot meaningfully be compared between patients.
h. Health-related quality of life measures would be desirable but none are well-established or universally accepted; they should not be used until clinically validated.
i. Global evaluation of study medication.
j. Pharmacoeconomic measures.
k. Incidence and nature of adverse events.

III.3.i.D. Study design

These are medium- or long-term studies (at least 4 months) in outpatients. At least two should be randomised, double-blind, placebo-controlled parallel-groups studies and at least one of these should include an active comparator. Three-arm trials, including placebo, are required for internal validation with active comparators unless the study is designed to show superiority over a well-established comparator (if superiority is not shown, non-inferiority cannot be claimed in the absence of placebo control). Randomisation should occur after a run-in (baseline) period of at least one month, when stratification for baseline attack rate (e.g., ≥3 or <3 per 4 weeks) is recommended as the prophylactic effect may depend on this variable. Treatment periods should be at least 3 months. Patients should take their usual acute therapy as required provided that it can be safely administered with the study drug. Attacks (and, if required, their features), acute medication use and adverse events should be recorded as they occur by patients in paper or electronic diaries. Reviews of patients and their diaries should be undertaken every 4 weeks, and should continue for at least 4 weeks after treatment is discontinued.

Compliance should be monitored.

III.3.i.E. Planned sample

Sample size calculations should be based on demonstrating superiority over placebo in a primary analysis of a) difference in attack frequencies, with a relative difference of 50% or an absolute difference of 1 attack/month being clinically significant and allowing for a reduction on placebo of up to 30% or 1 attack/month; or b) difference in responder rates, with an absolute difference of 20% being clinically significant and allowing for a placebo rate of up to 30%.

Whether or not greater numbers are required for safety evaluation is dependent on whether or not the study drug has been through a development programme, and is licensed already, for another indication.

III.3.i.F. Study population

a. Adults with frequent attacks of migraine with or without aura.
b. Adolescents and/or children, if they are to be included in the labelling.
III.3.i.G. Specific inclusion criteria
   a. Patients with migraine conforming to IHS diagnostic criteria 1.1 or 1.2 for at least 1 year and with at least 3 months’ well-documented retrospective history.
   b. Migraine attacks occurring 2-6 times monthly.
   c. Males and females.

III.3.i.H. Specific exclusion criteria
   a. Age at onset of migraine of 50 years or over.
   b. Other headaches not well distinguished from migraine or occurring with such frequency as to interfere with assessments (usually taken to mean occurring on >6 days/month).
   c. Other illnesses likely to interfere with assessments.
   d. Use of other migraine prophylactic drugs in the previous month.
   e. Use of or requirement for other unacceptable concomitant therapy.
   f. Risk of pregnancy.
   g. History of drug or alcohol overuse.

III.3.i.I. Tools for assessing endpoints
   Paper or electronic diaries.

III.3.i.J. Specific criteria for early withdrawal and discontinuation
   a. Withdrawal of consent (which may be related to lack of efficacy or adverse events).
   b. Medical concerns related to evident lack of efficacy or adverse events.
   c. Other intercurrent illness.
   d. Pregnancy.
   e. Breach of double-blinding.

III.3.i.K. Data analysis method
   Analysis should be based on the intention-to-treat (ITT) population. Because time to onset of effect is of interest, analysis of efficacy should be for the entire treatment period as well as for the period specified by the primary endpoint (e.g., the final 4 weeks of treatment). Standard statistical methods are appropriate. Adverse events are usually analysed descriptively.

   Cost-effectiveness (or cost-utility) analysis is highly desirable, but the methodology is not yet well developed.

   Longer-term studies are desirable to investigate continuing efficacy during and after periods of treatment (up to 6 months or longer) common in routine practice. At least one trial of at least 12 months’ duration is required for safety evaluation. These studies can be conducted as continuations of double-blind studies with patients opting, or not, to remain on their treatment (whether active or placebo). In addition, withdrawal of medication at the end of the prescribed period of treatment should be evaluated, ideally by randomised and double blind substitution of placebo in one group of patients and continuation of active therapy in another (with informed consent). Open observational studies, using patients as their own controls, have very limited value and are a poor alternative because of the inherent variability over time of the disease.

III.3.ii. Chronic tension-type headache prophylaxis

III.3.ii.A. Objectives
   To confirm efficacy, effectiveness and tolerability in treating chronic tension-type headache.

III.3.ii.B. Primary end-points
   a. Number of days with headache per specified unit time (usually 4 weeks) measured during treatment after a specified period (at least 8 weeks).
b. Response rate: percentage of patients with reduction in headache days per unit time of 50% or more (implying reversion from chronic to episodic tension-type headache) after a specified treatment period.

III.3.ii.C. Secondary endpoints
a. Number of days with headache over the entire treatment period.
b. Intensity of headache on a visual analogue scale or 4-point verbal rating scale [0 = no pain; 1, 2, 3 = mild, moderate, severe pain]) averaged over attacks within a specified evaluation period.
c. Duration of headache each day.
d. Headache indices multiplying frequency, intensity and/or duration are not recommended: arbitrary weighting in the numerical scores, which may be faulty, is increased by multiplication; and indices cannot meaningfully be compared between patients.
e. Functional measures and health-related quality of life measures would be desirable but are not established and should not be used until clinically validated.
f. Drug consumption for symptomatic or acute treatment totalled over an evaluation period.
g. Global evaluation of study medication.
h. Incidence and nature of adverse events.

III.3.ii.D. Study design
These are medium- or long-term studies (at least 4 months) in outpatients. At least two should be randomised, double-blind, placebo-controlled parallel-groups studies. There are no licensed active comparators. Randomisation should occur after a run-in (baseline) period of at least one month during which the number of days with headache and acute or symptomatic medication consumption are recorded. Stratification is unnecessary. Treatment periods should be at least 3 months. Days with headache, intensity and duration of headache, acute medication use and adverse events should be recorded as they occur by patients in paper or electronic diaries. Reviews of patients and their diaries should be undertaken every 4 weeks, and should continue for at least 4 weeks after treatment is discontinued.

Acute medication is inappropriate treatment for this disorder and should not be encouraged (regular use of acute or symptomatic medication on >2 days per week will put the diagnosis in question as this approaches the threshold for medication-overuse headache). Compliance should be monitored.

III.3.ii.E. Planned sample
Sample size calculations should be based on demonstrating superiority over placebo in a primary analysis of difference in days with headache. A change from baseline of ≥50% represents a clinically significant benefit of treatment, but the response to placebo has not been well documented (a reduction of up to 30% should be anticipated). Alternatively the primary analysis may be of difference in response rates. An absolute difference of 20% would be clinically significant. Again the response rate to placebo has not been well documented but up to 30% should be anticipated.

III.3.ii.F. Study population
Adults with chronic tension-type headache drawn from secondary or primary care or from the general population.

III.3.ii.G. Specific inclusion criteria
a. Patients with chronic tension-type headache conforming to IHS diagnostic criteria 2.3 for at least 3 months and with at least 3 months’ well-documented retrospective history.
b. Males and females.
c. Unless otherwise justified, patients should be over 18 years of age.
III.3.ii.H. Specific exclusion criteria
   a. Age at onset of chronic tension-type headache of 50 years or over.
   b. Other headaches, especially migraine, not well distinguished from tension-type headache or occurring with such frequency as to interfere with assessments.
   c. Other illnesses, particularly depression, likely to interfere with assessments.
   d. Use of other prophylactic drugs in the previous month.
   e. Use of acute or symptomatic medication for headache on an average of >2 days per week over the previous 2 months.
   f. Other history of drug or alcohol overuse.
   g. Use of or requirement for psychotropic medication or other unacceptable concomitant therapy.
   h. Risk of pregnancy.

III.3.ii.I. Tools for assessing endpoints
   Paper or electronic diaries.

III.3.ii.J. Specific criteria for early withdrawal and discontinuation
   a. Withdrawal of consent (which may be related to lack of efficacy or adverse events).
   b. Medical concerns related to evident lack of efficacy or adverse events.
   c. Other intercurrent illness.
   d. Pregnancy.
   e. Breach of double-blinding.

III.3.ii.K. Data analysis method
   Analysis should be based on the intention-to-treat (ITT) population. Because time to onset of effect is of interest, analysis of efficacy should be for the entire treatment period as well as for the period specified by the primary endpoint (e.g., the final 4 weeks of treatment). Standard statistical methods are appropriate. Adverse events are usually analysed descriptively.

   Longer-term studies are desirable to investigate continuing efficacy over longer periods of treatment that may be necessary in routine management (6 months or longer). At least one trial of at least 12 months’ duration is required for safety evaluation. These studies can be conducted as continuations of double-blind studies with patients opting, or not, to remain on their treatment (whether active or placebo). In addition, withdrawal of medication at the end of the prescribed period of treatment should be evaluated, ideally by randomised and double blind substitution of placebo in one group of patients and continuation of active therapy in another (with informed consent). Open observational studies, using patients as their own controls, are of very limited value and a poor alternative because of the inherent variability over time of the disease.

III.3.iii. Prophylaxis of episodic cluster headache

III.3.iii.A. Objectives
   To confirm efficacy and evaluate effectiveness and tolerability in terminating a cluster period or in reducing frequency, intensity and/or duration of continuing cluster headache attacks.

III.3.iii.B. Primary end-points
   a. Frequency of attacks per specified unit time (usually 1 week) measured during treatment after a specified period (to allow treatment effect to develop) following dosage-stabilisation.
   b. Remission rate: percentage of patients whose attacks have ceased after a specified treatment period.
The number of attacks should be recorded irrespective of their intensity or duration. An attack treated successfully with acute medication but with relapse within 1 hour counts as one attack.

III.3.iii.C. Secondary endpoints

- Frequency of attacks over the entire treatment period.
- Time to remission.
- Intensity of cluster headaches averaged over a specified evaluation period.
- Duration of cluster headaches summed or averaged over a specified evaluation period.
- Drug consumption for symptomatic or acute treatment totalled over an evaluation period.
- Health-related quality of life measures, and measures of functional disability over the treatment period, would be desirable but are not established and should not be used until clinically validated.
- Global evaluation of study medication.
- Effects after withdrawal of treatment.
- Incidence and nature of adverse events.

III.3.iii.D. Study design

These are short- or medium-term studies depending on the treatment effect observed in phase II. However, studies over >2 weeks cannot be conducted against placebo notwithstanding that high spontaneous remission rates confound trials that are not placebo-controlled. Therefore, superiority over an established comparator must be shown in one or more randomised parallel-groups studies. Whilst a number of reasonably effective potential comparator drugs exist, they are unlicensed for this indication, associated with toxicity and tend to be used in ways that make it very difficult to achieve double-blindness. Open studies are more acceptable with objective end-points (e.g., remission rate).

No run-in (baseline) period is needed. Stratification is recommended for time since onset of the cluster period (e.g., ≥2 or <2 weeks) and gender as each may influence the prophylactic effect or spontaneous remission rate. Treatment periods may need to incorporate dose-titration and, following dosage stabilisation, are defined by the times prescribed for the primary end-point or by the study objective if this calls for longer-term therapy. They are unlikely to exceed 3 months and safety evaluation must be conducted within this period unless safety has been demonstrated already in longer-term use of the drug for other indications.

Patients should take their usual acute therapy whenever cluster headache is of at least moderate intensity provided that it can be safely administered with the study drug. Attacks and their intensity and duration (and, if required, their associated features), acute medication use and adverse events should be recorded as they occur by patients in paper or electronic diaries. Reviews of patients and their diaries should be undertaken every week in short-term studies and at least monthly in medium-term studies.

Compliance should be monitored.

After remission of the cluster period, whether spontaneous or treatment-related, prophylactic medication is withdrawn. At least one trial should observe the consequences of withdrawal over up to several weeks, since these may include relapse.

III.3.iii.E. Planned sample

Sample size calculations should be based on demonstrating superiority over placebo in a primary analysis of:

- a) difference in attack frequencies, with a relative difference of 50% being clinically significant and allowing for a reduction on placebo of up to 20%; or
- b) difference in remission rates, with an absolute difference of 20% being clinically significant and allowing for a placebo plus spontaneous resolution rate of up to 20%.
III.3.iii.F. Study population
a. Adults with episodic cluster headache.
b. Adolescents and/or children, if they are to be included in the labelling.

III.3.iii.G. Specific inclusion criteria
a. Patients with episodic cluster headache conforming to IHS diagnostic criteria 3.1 and in at least their second cluster period.
b. Any length of time from onset of the cluster period provided that its expected duration, from start of study medication, is greater than the treatment period specified by the primary end-point.
c. Acute attacks occurring between once every 2 days and 5 times per day.
d. Males and females.

III.3.iii.H. Specific exclusion criteria
a. Other headaches not well distinguished from cluster headache.
b. Other illnesses likely to interfere with assessments.
c. Other cluster headache prophylactic therapy in the previous week.
d. Use of or requirement for other unacceptable concomitant therapy.
e. Risk of pregnancy.
f. History of drug or alcohol overuse.

III.3.iii.I. Tools for assessing endpoints
Paper or electronic diaries.

III.3.iii.J. Specific criteria for early withdrawal and discontinuation
a. Withdrawal of consent (which may be related to lack of efficacy or adverse events).
b. Medical concerns related to evident lack of efficacy or adverse events.
c. Other intercurrent illness.
d. Pregnancy.

III.3.iii.K. Data analysis method
Analysis should be based on the intention-to-treat (ITT) population. Subgroup analysis for gender differences is recommended and should be specified a priori. Since time to onset of effect is of interest, analysis of efficacy should be for the entire treatment period as well as for the period specified by the primary endpoint. Standard statistical methods are appropriate. Adverse events are usually analysed descriptively.

IV. OTHER STUDIES (SPECIAL INDICATIONS AND PRAGMATIC STUDIES)

IV.1. Children and adolescents

Development of drugs for headache disorders in these age-groups is clearly required for two indications:
  a. acute treatment of migraine;
  b. prophylaxis of migraine.

Disease characteristics differ in migraine between children/adolescents and adults. The results of acute and prophylactic treatment studies in adults with migraine cannot be extrapolated to younger age-groups. Separate trials in children/adolescents are not required by regulators for initial marketing authorisation, but these age-groups will be excluded from product-labelling if sufficient efficacy and safety data are not included in the regulatory submission. It is not certain whether such differences exist in episodic and chronic tension-type headache, whilst cluster headache is very rare (although not unknown) in children. It is likely, however, that regulators will adopt the same approach in these disorders whilst markets may not be commercially viable.
Further dose-finding studies in these age-groups may be needed. During phase III, drugs with clear efficacy and safety in adults with migraine may be assessed in separate placebo-controlled trials for effectiveness and safety in children and/or adolescents. Pivotal studies will be large multicentre randomised double-blind trials incorporating one or more doses of study drug. They may use parallel-groups or (multiple) cross-over designs, with regulators strongly favouring the former. One or more studies should include an active comparator where licensed comparators exist. Three-arm trials, including placebo, are required for internal validation with active comparators unless the study is designed to show superiority over a well-established comparator (if superiority is not shown, non-inferiority cannot be claimed in the absence of placebo control).

Objectives, end-points, study designs, sample sizes, inclusion/exclusion criteria (other than age), tools for assessing end-points and data analysis methods are all generally similar to those in adult migraine trials. There are a few exceptions:

a. migraine attacks are usually shorter-lasting, so rapid efficacy is more important;

b. associated symptoms of nausea and vomiting are commonly more pronounced, and effective treatment of these may be a higher priority;

c. children with headache are likely to be put to bed to sleep (which is curative), so frequent assessments over several hours is often impractical;

d. prophylactic medication in children is inappropriate before a review has been conducted of lifestyle and possible triggers, which should be built into the protocol as part of baseline evaluation.

IV.2. The elderly

All primary headache disorders become significantly less prevalent after the age of 60 years. There are no special requirements for trials in the elderly, who should not generally be excluded from adult trials (subject to other inclusion/exclusion criteria). Elderly patients with migraine should have been suffering from this disorder for many years: onset of migraine over the age of 50 years is uncommon and predictive of symptomatic disease, which must be excluded.

IV.3. Menstrual migraine

Migraine in women may be hormonally-triggered and occur solely in close temporal relationship to menstrual periods (menstrual migraine) or it may be more loosely associated with menstruation with a tendency to occur at or around the time of periods (menstrually-associated migraine). It is unlikely that treatment of menstrually-associated migraine should differ from that of migraine generally, whereas other possibilities arise for the treatment of menstrual migraine.

The EMEA advises that studies in menstrual migraine, to be undertaken once efficacy and safety have been demonstrated in non-menstrual migraine, have in principle the same design and end-points as studies in non-menstrual migraine. Subgroups of patients with menstrual migraine included in several studies may be combined in a meta-analysis planned a priori. The temporal relationship between menses and migraine attacks should be stringently recorded, and for diagnostic purposes this is necessary for three cycles before trial entry. In acute treatment trials, an important secondary end-point is the percentage of patients pain-free at 2 hours and remaining pain-free, without use of further medication, for 48 hours after treatment. In prophylactic trials there is the option for monthly short-term perimenstrual prophylactic treatment. If this is being evaluated, the trial design should require continuous observation throughout each month since the possibility exists that attacks are merely postponed to later in the menstrual cycle.
IV.4. “Mild” migraine
The previously widely-adopted primary end-point for acute migraine treatment trials (“headache relief”) required that treatment was delayed until pain was moderate or severe. This is counter-intuitive and possibly counter-productive, and many patients will not do it routinely. Although it is in part justified by the argument that earlier treatment results in the inappropriate use of migraine-specific therapy for non-migraine headache, particularly episodic tension-type headache, this argument is not clearly evidence-based. There are good reasons for conducting additional trials of early treatment, whilst pain is still mild. The recommended primary end-point, “pain-free” rate (percentage of patients pain-free at 2 hours after treatment), can and should still be used. Secondary endpoints should include rate of relapse and percentage of patients pain-free at 2 hours and remaining pain-free, without use of further medication, for 48 hours after treatment.

IV.5. Acute treatment of migraine in the aura phase
Patients with migraine with aura have the opportunity to treat during the aura phase, before headache commences. Efficacy of treatment in this phase cannot be assumed from studies of treatment taken later in the attack. Separate studies are required, with similar endpoints.

IV.6. Pragmatic studies
Pragmatic trials attempt to replicate routine practice in the use of a drug rather than the conditions of a controlled experiment. They rarely support marketing authorisation applications but can usefully inform prescribing practitioners.

A major concern in acute migraine trials is that recommended end-points, chosen because they are relatively objectively measurable and have proved statistically robust in differentiating between active treatments and placebo, do not well reflect patients’ views of what they want from a treatment. One suggested design for a preference study dispenses to each patient a quantity of each of two or more comparator drugs (which, if blinded, can include placebo). The patient chooses which to use on the basis of accumulating personal experience of each. The rate of use of each is an index of preference. Other measures of “satisfaction” are needed also since preference for one treatment over another does not indicate that either is adequate.

No studies have yet compared acute migraine therapy alone with acute plus prophylactic therapy, but these are needed. End-points are likely to reflect quality-of-life or pharmacoeconomic measures.

V. STUDIES FOR THE REGISTRATION OF GENERIC DRUGS
Intense discussion is underway that should clarify the regulatory requirements for generic marketing, at least in Europe. The central issue is at what point generic manufacturers are entitled to make reference to an innovator’s clinical trials data to support a marketing authorisation application for a copy product. This issue will soon come to the fore in acute migraine therapy.

VI. EXAMPLES OF LANDMARK WELL-DESIGNED TRIALS


VII. SUGGESTED READING


211
Chapter 18. Alzheimer’s Disease and Other Dementias

Ettore Beghi, M.D. ¹

Giancarlo Logroscino, M.D., Ph.D. ²

Amos Korczyn, M.D., M.Sc. ³

¹ Department of Neurology
University of Milano-Bicocca
Monza
ITALY
and
Institute for Pharmacological Research “Mario Negri”
Milano
ITALY

² Department of Epidemiology
HSPH, Harvard University
Boston, MA
USA

³ Department of Neurology
Tel Aviv University Medical School
Tel Aviv
ISRAEL
I. INTRODUCTORY REMARKS

Disease definition, causes, and frequency
The term *dementia* refers to a syndrome characterized by progressive loss of cognitive functions, including memory and language, and changes in personality and behavior. The most common cause of dementia in the elderly is Alzheimer’s disease (AD) (about 60%), followed by vascular dementia (VaD) (20%). However, considerable overlap exists between AD and VaD, and these two disorders may be extremes of a spectrum. In particular, the clinical differentiation between these disorders is difficult. AD is characterized by the presence of major impairment in learning and in retaining new information and at least one of the following: impairment of complex tasks, impaired reasoning ability, impaired spatial ability and orientation, and impaired language. In VaD loss of cognitive functions roughly overlaps that of AD, except for the onset or worsening of the symptoms within three months of a stroke and/or presence on neuroimaging of brain infarctions involving cortical or subcortical structures, including white matter. Dementia must be distinguished from *mild cognitive impairment* (MCI), which is a new memory complaint, preferably corroborated by an informant, with objective evidence of impairment of short-term memory, all other cognitive functions being normal, and no substantial interference with daily living activities. MCI is frequently an intermediate stage between normal cognition and dementia and is a risk factor for dementia.

Dementing disorders are followed by an increasing disability resulting in social and occupational decline. Dementia may also occur in younger persons where the etiology includes brain trauma, schizophrenia, multiple sclerosis and AIDS. In underdeveloped countries other infections of the CNS may also be a common cause of cognitive decline.

Over the past two decades, a series of well-conducted epidemiological studies have shown that dementia is a common condition affecting up to 5% of individuals over 65 years in industrialized countries. Most population-based studies indicate a prevalence increasing with age from 1.5/100 between 65 and 69 years to over 30/100 at age 85-89. It is not clear if after age 80-85 the prevalence of the disease is still rising or it tends to stabilize. The incidence of dementia is about 1/100 per year after age 65. In 1997 there were about 2.3 millions individuals with AD in the U.S. (range 1.1 to 4.6 millions) where the number of new cases is about 360,000 each year. Due to the aging of the population, the prevalence of AD in the US will grow about 4-fold within the next 50 years if effective interventions to delay the onset of disease are not developed.

Goals of treatment
There are several goals of the treatment of dementia, including: 1) prevention of disease occurrence; 2) symptomatic improvement; 3) cure of disease; 4) delaying cognitive decline; 5) complete symptomatic control. At present there is no proven treatment that modifies the natural history of the disease or changes its outcome. Potential areas for intervention include: 1) increasing the levels of neurotransmitters involved in cognitive functions; 2) providing neuroprotection to neurons already damaged or showing functional changes; and 3) neuronal regeneration by replacing neurons which have been lost. In the future the goal will be to reverse the natural history of dementia by reversing the typical neuropathological lesions (β-amyloid and tau).

Symptomatic treatment of dementia
The first paper describing an effective symptomatic treatment of dementia was published less than twenty years ago and was a small cross-over study with tacrine, the first cholinesterase inhibitor (ChEI). In the last two decades, several drugs have been tested and approved. Patients with dementia of any severity treated in randomized clinical trials with donepezil, rivastigmine, and galantamine (which represent the main category of ChEIs shown to be effective with an acceptable tolerability profile) experience some benefits in cognitive function, activities of daily living and behavior at least during the first year of treatment. Memantine, an NMDA-receptor antagonist, and possibly ginkgo biloba have also been reported to be effective.
Problems with trials on symptomatic treatment of dementia

To the present time, the efficacy of symptomatic drugs is at best modest. However, the results of the published randomized trials must be interpreted in the light of methodological drawbacks, concerning especially the definition and the choice of the appropriate outcome. Many outcome measures have been used in dementia trials. The ideal outcome should be easy to measure and easily collectable at each follow-up over a significant period of time (ideally for several years). These measures should have a good reliability, especially considering that many trials are multicenter-based. Cognitive decline is difficult to measure with the available instruments in a quantitative way. Typical problems are floor and ceiling effects, regression to the mean, learning effects and placebo factors. Moreover the range of changes over a short period is small compared to the possible cognitive range of each scale. More suitable and robust end-points are the following: 1) loss of independence; 2) loss of a specific daily living function; 3) placement in a nursing home. However, these end-points require prolonged follow-up and may be influenced by other environmental factors like the presence of an active caregiver, and economic and social environment.

Cognitive tests include a battery of tests covering memory and other domains including language, constructional abilities, attention/concentration and psychomotor speed. Remote and recent memory must be extensively explored along with recall and recognition for various modalities. Verbal and visuo-spatial memory must be also investigated. The Mini Mental State Examination (MMSE) is generally used as a screening test for cognitive impairment while the Clinical Dementia Rating (CDR) is used for grading disease severity. Although the Alzheimer’s Disease Assessment Scale cognitive subscale (ADAS-cog) fulfils the requirements of a comprehensive cognitive scale, none of the available instruments can be preferred as being more valid and reliable. Although several scales have been proposed to measure the activities of daily living (ADL) and to assess an overall clinical improvement, none has advantages that would justify its preferred use in regulatory trials. Moreover, all these tests have shown acceptable validity and reliability in AD, but not in other dementing disorders. This is of particular concern in some aspects. For example, the MMSE is heavily weighted to memory and left hemisphere functions, e.g. speech.

Criteria for assessing symptomatic improvement

The Aging-Warner consortium established in 1992 that two types of outcomes should be used to assess the efficacy of an anti-dementia product: 1) a global assessment performed by a skilled clinician; 2) a performance-based objective test of cognitive function. The consortium identified the Clinician’s Global Impression Scale (CGI) and the Scale ADAS-cog as two preferable instruments. This two-outcome approach has the goal to identify changes that are at the same time clinically meaningful and specific (as determined by cognitive testing) and that are not due to some non-specific effects on the general clinical state. In line with the Note for Guidance on Medicinal Products of the European Agency for the Evaluation of Medicinal Products (EMEA), the efficacy criteria for a drug tested for the treatment of dementia may include symptomatic improvement (which may be manifest in enhanced cognition, more autonomy and/or improvement in behavioral dysfunction), slowing or arrest of symptom progression, and primary prevention of disease by intervention at a pre-symptomatic stage. The EMEA requires a six-month treatment for demonstration of efficacy and one year for maintenance of efficacy. Improvement of symptoms should be assessed in three specific domains: 1) cognition, as measured by neuropsychological tests (cognitive end-point); 2) activities of daily living (functional end-point); and 3) overall clinical response (global end-point). Efficacy variables should be defined for each domain. Primary efficacy variables should include cognitive end-points and the clinical relevance of the improvement in cognition, measured preferably by a functional end-point. The overall benefit should then be measured in terms of the proportion of patients achieving a meaningful benefit (responders). Other end-points include behavioral symptoms, which may be selected as secondary end-points or primary end-points in trials designed to assess control of behavioral abnormalities. The Food and Drug Administration (FDA: http://www.fda.org) has similar guidelines with a two-outcome approach to assess the efficacy of a drug (one cognitive and one global assessment measure). The FDA requires three-month duration with at minimum 1000 patients exposed for several weeks in the relevant dosage range.
Prevention of dementia
The methodological concerns with symptomatic treatment trials (e.g., cognitive function testing) also apply to prevention trials. RCTs for prevention require very large sample sizes to ensure enough power to detect a significant reduction in the incidence of dementia. Because of this problem, most of the data available to date are based on observational studies. A possible strong end-point could be the clinical diagnosis of dementia (incidence of disease) in previously non-demented subjects. A clinical diagnosis of MCI has more problems in terms of definition and reliability. An additional problem of prevention trials is the long duration of follow-up necessary. The selection of subgroups at higher risk can make it easier to reach the numbers needed. Higher risk groups could be subjects carrying APO-E4, subjects in older age groups, or subjects with MCI. However, results from studies in such selected groups may not necessarily be generalized to wider populations. The incidence of AD is influenced by several factors, such as age, sex, education, family history of disease and genetic status (APO-e), which may affect the outcome of a trial. In a recent study, Kriyscio and coworkers (1) revised the factors affecting the probability of AD in a preventive trial: length of the follow-up period, accrual, incidence rate of disease, drop-in, drop-out rate and the adherence rate. PREADVISE, which is the largest prevention trial up to date, focuses on assessing the potential protective effect of selenium and vitamin E and plans to enroll 10,700 subjects at 400 sites. Another factor that can increase sample size is the rate of misdiagnosis, especially for mild new cases that could not be identified with the operational diagnostic instruments of the trial.

The use of placebo in dementia trials
Given the great heterogeneity of dementing disorders in terms of symptom profile, overall severity and course, the efficacy of a drug can be demonstrated only by using appropriate controls. The availability of active treatments raises ethical concerns about the use of placebo in patients with dementia. However, use of placebo may still be considered for several reasons: 1) Drug efficacy could possibly be only detected with placebo as a comparator; because superiority against an active control may be difficult to prove. Equivalence or non-inferiority designs (requiring larger samples) may be used as proof of efficacy in studies using active controls, but such studies raise interpretative problems, because the argument could be raised that the reference control was not necessarily efficacious in the population and under the conditions being studied; 2) There is no evidence that available treatments affect the patient’s long-term health; 3) The efficacy of cholinesterase inhibitors is at best modest; 4) Lack of effective drugs for the prevention of dementia justifies the use of placebo to test drugs assessed for this indication; 5) There are difficulties in determining rates of adverse effects against active comparators which may also cause adverse effects.

The Quality Standard Subcommittee of the American Academy of Neurology (2) recognized that the use of cholinesterase inhibitors should be considered a standard of care for AD patients, even if the clinical effect is small. This requirement may justify placebo-controlled trials only when the investigational agent is used as add-on to existing treatments.

Issues with informed consent
Putative anti-dementia drugs typically undergo a staggered development process, culminating in double-blind studies, usually with a parallel design in which placebos are employed. Patients asked to participate in controlled trials of an investigational treatment must be informed of any alternative treatment and should be able to explore the positive and negative consequences of the treatments being tested (including the use of placebo) to provide a fully informed consent.

Participation of demented patients in such studies calls for attention to the special circumstances of this population and their needs, including consideration of potential benefits to the patients as well as caregivers, the economic impact, and the expected benefits to the society and science. However, demented patients may not be able to understand the full implications of the study, and may be unduly influenced by researchers and caregivers. In addition, the practice of obtaining proxy consent from the patient’s surrogate does not satisfactorily resolve the ethical issues, as the surrogate’s decisions usually reflect their
personal rather than the patient’s choice. In summary, the inability of patients to fully comprehend the possible implications of the study and the consequent need for the consent to be provided by a surrogate raise notable problems. The investigators and the IRB have an important role in ensuring that drug studies for patients with dementia are performed in a way that provides optimal information and preserves the well-being of patients as well as support for their caregivers.

**Biological markers**

Biomarkers can be defined as biological compounds that can be measured as indicators of exposure (risk factors of disease), intermediate steps of pathogenic pathways, or different clinical stages of the disease including specific responses to drug therapy. Genetic markers (e.g. the apolipoprotein E genotype) can help to identify subjects who are at higher risk of dementia. Several blood and CSF tests have been proposed for the early detection of AD. Such markers should reflect the pathophysiological mechanisms of AD, an example being the measurement of brain, serum or CSF β-amyloid and tau protein concentrations to detect altered metabolism of amyloid and neurofibrillary degeneration. Tau is a microtubule-associated protein that forms the basic element of the neurofibrillary tangle, one of the characteristic lesions of AD. CSF-tau levels have shown a good sensitivity (85%) and specificity (83%) to distinguish AD patients from normal elderly controls. Aβ42 is a 42 aminoacid fragment of the transmembrane amyloid precursor protein (APP) that aggregates as β-pleated sheets in extracellular neuritic plaques. Aβ42 and a shorter 40-amino-acid peptide (Aβ40) can both be assayed in the CSF. Several studies have consistently demonstrated a moderate to marked decrease in CSF Aβ42 in AD, probably because this compound is bound within the neuritic plaques.

At present, no biological marker has been recommended for use in clinical trials of dementia. The use of biological markers could be useful in the future, especially among subjects who are asymptomatic or have MCI, to select subgroups at higher risk to develop dementia. Alternatively, biological markers can be considered as surrogate end-points to assess treatment efficacy. A consensus committee has recently proposed a reclassification of biomarkers for AD in clinical practice (3). These include core markers (those judged to have reasonable evidence for association with key mechanisms of AD pathology) and non-core markers (those felt to be less clearly associated with mechanisms of pathogenesis or neurodegeneration in AD). Core markers include amyloid beta peptide, APP, tau proteins, isoprostanes, A1-antichymotrypsin, interleukin-6-receptor-complex, C-reactive protein, C1q, homocysteine, oxysteroids, 3-nitrotyrosine. Non-core markers include glutamine synthetase, human antibodies against Aβ-related proteins, glial fibrillary acidic protein, sulfatide, AD7C/NTP, and kallikrein 6.

**II. PHASE II STUDIES FOR REGISTRATION OF NEW SYMPTOMATIC DRUGS**

**II.1. Outline of a typical development plan**

During this phase the candidate drug is tested against placebo. The goal of this phase is to document efficacy and to identify the parameters of the treatment regimen (titration, dose regimen, maximal tolerated dose, etc.) most likely to maximize the therapeutic response in patients with well-defined disease. As a rule, phase II studies are designed to maximize efficacy by using the smallest possible number of patients with homogeneous disease characteristics, notably those with fewer concomitant illnesses and less severe impairment, in whom clinical response can be detected over a relatively short time. To increase sensitivity, drug response may also be tested by enrolling patients who responded during a pre-randomization phase. Phase II controlled trials must be preceded by open label exploratory studies to assess titration rates, maximally tolerated doses and pharmacokinetics.

The typical trial design is randomized, placebo-controlled, parallel group testing at least two dose regimens over a short time period. Titration to the predetermined doses should be identified to minimize
drop-outs for adverse effects. Patients included in short term phase II clinical trials should be allowed to participate in long-term trials.

II.2. Short-term phase II studies

II.2.A. Objectives
To evaluate short-term efficacy and tolerability and to detect a correlation between different doses and positive and untoward effects.

II.2.B. Primary end-points
1. Change in cognitive function (measured by psychometric tests)
2. Clinical global impression of change
3. Change in performance of ADL
4. Acceptability of treatment as measured by withdrawal from trial
5. Safety as measured by incidence of adverse events, particularly those leading to withdrawal

II.2.C. Secondary end-points
1. Behavioral disturbances
2. Change in quality of life
3. Effect on caregiver

II.2.D. Exploratory end-points
1. Plasma drug levels
2. Changes in functional imaging
3. Effects on biological markers

II.2.E. Study design
Multicenter, randomized, placebo-controlled, parallel group. A cross-over design may be employed in short-lasting treatment periods and when carry-over effects are insignificant. A screening phase may be used to verify patient eligibility, followed by a prospective baseline period during which cognitive functions are tested and the functional and global clinical activities are measured. After randomization, a titration period is started of sufficient length to achieve steady state conditions. A maintenance period of six months follows under the assumption that during this period there will be a clinically significant progression of the disease (for example, a 4-point change on the ADAS-cog) in the placebo arm. Eligible patients should be free of concomitant illnesses and taking no or few active principles. At the end of the trial, the patient is either withdrawn according to a pre-defined treatment schedule, or he/she enters a long-term phase.

II.2.F. Planned sample
With two treatment arms (active vs placebo), a sample size of about 100 per treatment group is needed with an expected 15% of responders (e.g. a 4-point or greater improvement on the ADAS-cog) in the placebo group, under the assumption to detect a 20% absolute difference in the proportion of responders, with an 80% power, a 5% (two-sided) significance, and a 20% drop-out rate.

II.2.G. Study population
Patients with definite dementia, AD or VaD.

II.2.H. Specific inclusion criteria
a. Adult female and male patients
b. MMSE between 10 and 26
c. Reliable caregiver
d. Dementia of mild to moderate severity (CDR<3 within 4 weeks prior to entry).
e. Imaging studies performed during six months prior to entry consistent with the diagnosis.

II.2.I. Specific exclusion criteria
   a. Delirium or impairment of consciousness
   b. Major depression or other significant psychiatric diagnosis
   c. History of drug or alcohol abuse
   d. Other disorders possibly causing dementia
   e. History of hypersensitivity to relevant drugs.
   f. Neoplastic, hepatic, renal or cardiac disorders of significant impact on function or survival
   g. Any disability preventing compliance with test procedures.

II.2.J. Tools for assessing primary end-points
   a. ADAS-cog or other valid and reliable cognitive scale;
   b. CGIC, CIBIC plus or other valid and reliable scale assessing overall clinical impairment

II.2.K. Tools for assessing secondary end-points
   a. PDS, IADL or other valid and reliable scale assessing ADL, quality of life, cognitive and behavioral abnormalities
   b. Caregiver global impression
   c. Nurse global impression

II.2.L. Specific criteria for early withdrawal and discontinuation
   a. Occurrence of significant adverse events thought to impair ADL and overall quality of life
   b. Poor compliance
   c. Withdrawal of consent

II.2.M. Data analysis method
The analysis of treatment efficacy is performed on the intention-to-treat population (all randomized patients receiving at least one dose of study medication). Parametric and non-parametric tests are used as appropriate for primary and secondary end-points. Continuous variables (cognitive scores) are tested using parametric tests, like the Student’s t test or analysis of variance, and non-parametric tests, like the Wilcoxon-Mann Whitney test. Categoric variables (global impression, ADL, IADL, proportion of responders or cases withdrawn from the study, etc.) are tested using the chi-square test (parametric) or the Kruskal-Wallis test (non-parametric). Changes in the rate of decline can be assessed with survival analysis (Kaplan-Meier survival curves and Cox’s proportional hazard function). Univariate and multivariate statistical techniques can be used as appropriate. All p values should be based on two-sided tests with a 5% significance level.

II.3. Long-term phase II studies
When treatment efficacy is demonstrated in short-term clinical trials, long-term phase II studies are implemented to verify the duration of treatment effects on cognitive, functional and behavioral parameters. The estimated duration of a long-term clinical trial is about 12 months. During this period safety and efficacy are investigated with respect to symptom relief, slowing of progression of cognitive decline, and control of behavioral abnormalities. Ideally, long-term clinical trials are the extension of short-term studies. The dosage of the study medication may be adjusted to achieve the maximally tolerated dose. At study end all patients, including those who were in the placebo arm, should be given the active medication, which should be continued as long as the physician and/or caregiver perceives it to be beneficial, and retention time should be used as a measure of treatment effectiveness and tolerability. Along with cognitive tests, treatment benefits should be measured in terms of effects on hard end-points like time to loss of independence and/or relevant functional impairment.
II.3.A. Objectives
To evaluate long-term efficacy and tolerability of treatment

II.3.B. Primary end-points
a. Change in cognitive function (measured by psychometric tests)
b. Clinical global impression of change
c. Change in performance of ADL
d. Time to loss of independence and/or relevant functional impairment
e. Retention time as a measure of treatment efficacy
f. Acceptability of treatment as measured by withdrawal from trial
g. Safety as measured by incidence of adverse events leading to withdrawal

II.3.C. Secondary end-points
a. Behavioral disturbances
b. Change in quality of life
c. Effect on caregiver

II.3.D. Exploratory end-points
Effects on biological markers

II.3.E. Study design
Multicenter, randomized, placebo-controlled, parallel group.

II.3.F. Planned sample
Sample size should be calculated on end-points like loss of independence (placement in nursing home) or severe functional impairment; given a 20% expected 12-month rate in the control group, a sample of about 400 per treatment group is needed under the assumption to detect a 10% difference in the proportion of patients achieving the end-point, with a 80% power, a 5% (two-sided) significance, and a 25% drop-out rate.

II.3.G. Study population
As indicated for short-term phase II studies

II.3.H. Specific inclusion criteria
As indicated for short-term phase II studies

II.3.I. Specific exclusion criteria
As indicated for short-term phase II studies

II.3.J. Tools for assessing primary end-points
As indicated for short-term phase II studies

II.3.K. Tools for assessing secondary end-points
As indicated for short-term phase II studies

II.3.L. Specific criteria for early withdrawal and discontinuation
As indicated for short-term phase II studies

II.3.M. Data analysis method
As indicated for short-term phase II studies
III. PHASE III STUDIES FOR REGISTRATION OF NEW SYMPTOMATIC DRUGS

III.1. Outline of a typical development plan

During phase III development at least one large multicenter, randomized, double-blind, placebo-controlled, parallel-group confirmatory trial must be undertaken. A factorial design can also be considered comparing the investigational drug with cholinesterase inhibitors and placebo. Patients to be enrolled should have a definite diagnosis, like probable AD or VaD. In contrast to phase II trials, an effort should be made to include patients representative of those usually seen in clinical practice. As well, the choice of daily drug doses and dose increments should follow the patterns of clinical practice. As with phase II studies, outcome measures must include changes in cognitive scores and functional and behavioral changes. As the goal of phase III studies is to provide evidence of sustained treatment efficacy over a prolonged time period, the estimated length of the study should be about 12 months during which a statistically significant and clinically relevant difference should be documented in favor of the experimental treatment. In studies employing a factorial design, an additive effect is also searched in favor of patients receiving the investigational drug and the other active treatment. Having tested the daily dosage with the best therapeutic ratio in phase II trials, in these pivotal studies the drug should be titrated upwards to reach the highest tolerated dose. At the end of the double-blind phase, an open label extension period should also be considered to test the drug over an even longer period of time and under conditions most likely to reproduce the setting of clinical practice.

III.2. Typical phase III study

III.2.A. Objectives
To evaluate sustained efficacy and tolerability in a sample population representative of that seen in clinical practice

III.2.B. Primary end-points
a. Change in cognitive function (measured by psychometric tests)
b. Clinical global impression of change
c. Change in performance of ADL
d. Time to loss of independence and/or relevant functional impairment
e. Retention time as a measure of treatment efficacy
f. Acceptability of treatment as measured by withdrawal from trial
  g. Safety as measured by incidence of adverse events leading to withdrawal

III.2.C. Secondary end-points
a. Behavioral disturbance
b. Change in quality of life
c. Effect on caregiver

III.2.D. Exploratory end-points
Effects on biological markers

III.2.E. Study design
Multicenter, double-blind, placebo-controlled, parallel group with or without factorial design

III.2.F. Planned sample
As with long-term phase II studies.

III.2.G. Study population
Patients with dementia, including probable AD and VaD.
III.2.H. Specific inclusion criteria
   a. Adult female and male patients
   b. MMSE between 10 and 26
   c. Reliable caregiver
   d. Dementia of any severity supported by imaging studies performed during six months prior to entry

III.2.I. Specific exclusion criteria
   As indicated for short-term phase II studies

III.2.J. Tools for assessing primary end-points
   As indicated for short-term phase II studies

III.2.K. Tools for assessing secondary end-points
   As indicated for short-term phase II studies

III.2.L. Specific criteria for early withdrawal and discontinuation
   As indicated for short-term phase II studies

III.2.M. Data analysis method
   As indicated for short-term phase II studies.

IV. OTHER STUDIES

IV.1. PREVENTION TRIALS IN MILD COGNITIVE IMPAIRMENT (MCI)

IV.1.A. Outline of a developmental plan

As patients with MCI are expected to convert to dementia at a rate of about 10-15% per year, they represent the ideal target for a prevention trial. Using as an end-point the conversion to dementia diagnosed according to the DSM-IV or NINCDS-ADRDA criteria, a three-year trial has sufficient length to document a statistically significant, clinically relevant, and sustained treatment effect. Placebo must be used to detect the effects on disease progression attributable to active treatment. This procedure is however not without problems. MCI patients who convert to AD within 3 years are very likely to have the disease already at baseline, and thus the study is not really designed for prevention of dementia but rather examines the effect of the drug on the rate of cognitive decline. A practical issue is that MCI patients who are recruited do not necessarily develop dementia at the high rate reported in the literature, and may not represent the real world of MCI. In particular, these subjects may have a more benign course because they have a more prolonged course or have higher prevalence of anxiety or depression.

IV.1.B. Representative trial protocol

IV.1.C.i. Objectives
   To assess the treatment effects on disease progression and conversion to dementia

IV.1.C.ii. Primary end-point
   Time to the diagnosis of dementia.

IV.1.C.iii. Secondary end-points
   Time to the diagnosis of AD, VaD, and other dementia types.
IV.1.C.iv. Exploratory end-points
Effects on biological markers.

IV.1.C.v. Study design
Multicenter, double-blind, placebo-controlled, parallel group with or without factorial design.

IV. 1.C.vi. Planned sample
Given the expected 15% 12-month conversion rate on which the treatment response is measured, a sample of about 120 subjects per treatment group is needed under the assumption to detect a 20% in the proportion of responders, with an 80% power, a 5% (two-sided) significance, and a 30% drop-out rate.

IV. 1.C.vii. Study population
Patients with MCI, i.e. individuals with a new deficit in at least one cognitive domain (usually recent memory) but who appear to function independently in ADL.

IV. 1.C.viii. Specific inclusion criteria
a. Adult female and male patients
b. Presence of a memory complaint, preferably corroborated by an informant
c. Objective evidence of impairment of short-term memory (for age)
d. Otherwise normal cognitive functions
e. No interference with work, social activities, or other ADL
f. MMSE > 26 (Absence of dementia)

IV.1.C.ix. Specific exclusion criteria
a. Major depression, anxiety or other significant psychiatric diagnosis
b. History of drug or alcohol abuse
c. Other disorders possibly causing cognitive decline
d. History of hypersensitivity to relevant drugs
e. History of (active) neoplasm, hepatic, renal or cardiac disorders, which could affect the patient’s survival
f. Any disability preventing compliance with test procedures

IV.1.C.x. Tools for assessing primary end-points
a. NINCDS-ADRDA criteria for the diagnosis of probable dementia

IV.1.C.xi. Tools for assessing secondary end-points
b. DSM-IV criteria for the diagnosis of dementia
c. NINDS-AIREN criteria for the diagnosis of vascular dementia, Hachinski Ischemic Scale for the assessment of vascular dementia
d. Work Group on Frontotemporal Dementia and Pick’s Disease diagnostic criteria
e. Consensus Guidelines for the diagnosis of dementia with Lewy bodies

IV.1.C.xii. Specific criteria for early withdrawal and discontinuation
a. Occurrence of serious adverse events or events thought to impair ADL and quality of life
b. Poor compliance
c. Withdrawal of consent.

IV.1.C.xiii. Data analysis method
The analysis of treatment efficacy is performed on the intention-to-treat population. Univariate and multivariate statistical techniques can be used as appropriate (see also short-term phase II studies).
Conversion to dementia (in general and by type) can be assessed with survival analysis (Kaplan-Meier survival curves and Cox’s proportional hazard function). All p values are two-sided with a 5% significance level.

**IV.2. PREVENTION TRIALS IN ASYMPTOMATIC PATIENTS**

**IV.2.1. Outline of a developmental plan**

A trial on the prevention of dementia in asymptomatic elderly individuals must be designed considering the expected incidence of dementia in the study population and the factors most likely to affect the incidence of the disease. These factors include age, family history of disease, race/ethnicity, education, and genetic background. In addition, the expected number of patients developing dementia depends on the length of the follow-up period and the drop-out rate (mostly caused by death and poor compliance). Several assumptions are thus required for the calculation of the accrual period, the estimate of the sample size, and the duration of the follow-up.

**IV.2.2. Representative trial protocol**

**IV.2.A.i. Objectives**
To assess the treatment effects on the incidence of dementia.

**IV.2.A.ii. Primary end-point**
Reduction of the incidence of dementia.

**IV.2.A.iii. Secondary end-points**
Reduction in the incidence of dementia in patient subgroups defined by age, education, and genetic factors.

**IV.2.A.iv. Exploratory end-points**
Effects on biological markers.

**IV.2.A.v. Study design**
Multicenter, double-blind, placebo-controlled, parallel group with or without factorial design.

**IV.2.A.vi. Planned sample**
Based on the calculations made for the PREADVISE study, a sample of about 2700 individuals per treatment group is required to halve the incidence of dementia with a 80% power, a 5% level of significance, and a 30% drop-out rate.

**IV.2.A.vii. Study population**
Asymptomatic elderly individuals.

**IV.2.A.viii. Specific inclusion criteria**
- a. Female and male patients aged 65 years and older
- b. Normal cognitive functions

**IV.2.A.ix. Specific exclusion criteria**
- a. Major depression, anxiety or other significant psychiatric diagnosis
- b. History of drug or alcohol abuse
- c. Other disorder possibly causing cognitive decline
- d. History of hypersensitivity to relevant drugs
e. History of (active) neoplasm, hepatic, renal or cardiac disorders, which could affect the patient’s survival
f. Any disability preventing compliance with test procedures

IV.2.A.x. Tools for assessing primary end-points
NINCDS-ADRDA criteria for the diagnosis of probable dementia

IV.2.A.xi. Tools for assessing secondary end-points
Genetic background can be defined by history taking and APO-E genotype

IV.2.A.xii. Specific criteria for early withdrawal and discontinuation
a. Occurrence of serious adverse events or events thought to impair ADL and quality of life
b. Poor compliance
c. Withdrawal of consent.

IV.2.A.xiii. Data analysis method
The analysis of treatment efficacy is performed on the intention-to-treat population. Univariate and multivariate statistical techniques can be used as appropriate (see also short-term phase II studies). Incidence of dementia (in general and by type) can be assessed with survival analysis (Kaplan-Meier survival curves and Cox’s proportional hazard function). All p values are two-sided with a 5% significance level.

IV.3. PRAGMATIC TRIALS
As in other clinical conditions, pragmatic trials are designed to reproduce settings reflecting more closely the use of a drug in clinical practice. Pragmatic trials may be designed to assess the effectiveness of different therapeutic strategies (e.g., early vs delayed treatment) and to test treatment in populations usually not included in regulatory trials (e.g., oldest patients or patients with concurrent disabling disorders). Studies comparing different drugs and allowing dosing flexibility could be considered. Survival analysis with retention time as the primary end-point should be the preferred choice for measuring treatment effectiveness. Other outcome measures could include time to nursing home placement or loss of independence. An intent-to-treat analysis should be performed in all cases. Trial duration may vary according to the type of therapeutic strategy but it should be generally no shorter than 24 months.

IV.4. SPECIAL INDICATIONS
The designs described for phase III clinical trials can also be used for the assessment of efficacy in patients with specific syndromes (dementia associated with cerebrovascular disorders, dementia with Lewy bodies, fronto-temporal dementia). In these cases, the inclusion/exclusion criteria, the primary and secondary end-points, and the relative tools are the same as those used for dementia at large. As with the management of concurrent clinical conditions, the treatment of the underlying disorders (stroke, Parkinsonism, etc.) should be carefully considered in terms of interactions and specific contraindications.

IV.4.i. Studies in Vascular Dementia
Vascular dementia (VaD is considered the second most common form of dementia after AD worldwide but probably the first in some countries. Cerebrovascular disease can determine VaD with different mechanisms including large-vessel disease with multiple strokes, single strokes in strategic areas, or subcortical lesions with multiple lacunar infaracts and white matter lesions. The diagnosis of VaD is possible when dementia, history of cerebrovascular disease and a relationship between the two disorders is present. The characteristic feature of subcortical VaD is the involvement of executive functions. These include ability to execute complex behaviors and solving-problems ability. MMSE is not a good
instrument to assess executive functions. Several tests (among which the Trail-Making test or the Clock Drawing task) can assess executive functions. These tests should be included in the assessment of the diagnosis and follow-up of VaD.

Several risk factors are associated with cerebrovascular diseases and consequently with VaD.

The prevention of cerebrovascular disease should be the first step in the prevention of VaD. Studies looking at the efficacy of controlling cardiovascular risk factors in the prevention of VaD are few. In the non-demented subjects enrolled in the Syst-Eur study who received antihypertensive treatment the risk of dementia was less than 50% of that of controls (4).

There is growing evidence that in VaD as in AD there is involvement of the cholinergic system. Animal models of stroke-prone spontaneously hypertensive rats present behavior that can be considered similar to memory impairment present in VaD. These rats show reduction of acetylcholine in several areas of the brain including hippocampus. Cholinergic agents have therefore been tested as potential treatments in VaD. In the largest trial, 603 subjects were recruited for a multicenter randomized trial on donepezil in VaD (5). Patients with probable and possible VaD were recruited. Donepezil was found to be effective for patients with VaD using the ADAS-cog and CIBIC–plus as outcome measures.

IV.4.ii. Behavior and mood disturbances in demented subjects

Depression, anxiety, agitation and more serious symptoms like delusions and aggressive behavior are commonly seen in AD. Dementia-related behavioral disturbances have been associated with excess disability, increased caregiver burden, and premature institutionalization. The presence of behavioral disturbances is one of the main reasons for exclusion of patients from clinical trials. Pharmacotherapy is often necessary to treat these disturbances and specific trials are indicated to document efficacy and tolerability for this specific indication. For example, a NIH-sponsored trial is currently recruiting AD patients to study the effects of citalopram and risperidone in people with dementia-related behavior problems.

V. EXAMPLES OF LANDMARK WELL-DESIGNED TRIALS

postmenopausal women: the Women's Health Initiative Memory Study: a randomized controlled trial. JAMA 2003;289:2651-2662.

VI. SUGGESTED READINGS


VII. REFERENCES

Chapter 19. Parkinson’s Disease and Other Extrapyramidal Disorders

Caroline M. Tanner M.D., Ph.D. 1

Christopher G. Goetz, M.D. 2

Yoav Ben-Shlomo, MBBS, Ph.D. 3

Piu Chan, M.D., Ph.D. 4

1 Parkinson’s Institute
1170 Morse Avenue
Sunnyvale, CA 94089
USA
Tel: 408-734-2800
Fax: 408-734-8455

2 Rush Medical Center
Chicago, IL
USA

3 Department of Social Medicine
University of Bristol
Bristol
UK

4 Xuan Wu Hospital of Capital University of Medical Sciences
Beijing
CHINA

Corresponding Author:
Caroline M. Tanner M.D., Ph.D.
I. INTRODUCTORY REMARKS

Parkinson’s disease (PD) is the second most common neurodegenerative disorder of later life. The primary areas of neuropathologic injury are pigmented aminergic neuronal populations - the substantia nigra, locus ceruleus and others. PD causes progressive disability. Disability associated with PD includes not only motor dysfunction, but for many dysautonomia, cognitive changes (ranging from loss of executive function to frank dementia) and depression. There is no known cure nor is there a recognized method for slowing the ongoing degenerative process.

I.1. Unfulfilled therapeutic needs

Currently available therapies are most effective in minimizing the motor dysfunction of PD, through increasing the activity of the nigrostriatal dopamine system, although overall benefit is partial, and not sustained over the years of disease. Development of a treatment with sustained therapeutic benefit over the decades of the disease would be invaluable. In addition, these standard therapies also have disabling and/or dose limiting adverse effects. Acute adverse effects include nausea and hypotension. Chronic adverse effects include dyskinesias (involuntary movements), somnolence, hallucinations or psychosis. In more advanced disease, a progressive diminution in the overall motor benefit, and in the duration of the treatment effect is an additional limitation of some therapeutic approaches. Methods for preventing the development of these disabling adverse effects, or for treating them effectively once developed, are needed. Treatments for the nonmotoric disabilities of PD (such as dysautonomia, cognitive changes and depression) are few, and none are universally effective. Developing effective therapies for these nonmotor features of PD is an emerging area of interest.

Much effort has been directed to the development of treatments that can stop or slow the progression of established disease (neuroprotection), but so far such an agent has not been identified unequivocally. Similarly, no cure is available, and there is no way to delay the onset of PD. These latter questions are the most critical, since they address preventive or curative goals, rather than symptom amelioration.

Other related areas of importance involve pharmacogenetic investigations in PD and developing treatments for “atypical” parkinsonism. Little work has been done to determine whether the genetic makeup of the individual in part determines response to therapy. This would include investigating subgroups with genetic forms of parkinsonism, as well as PD patients with genetic variants of elements of the dopaminergic system (such as metabolic enzymes or receptors). The “atypical” parkinsonian syndromes include less common chronic neurodegenerative disorders with prominent parkinsonian features, such as multiple system atrophy and progressive supranuclear palsy. In contrast to PD, there are no effective therapies for these devastating disorders, although existing antiparkinsonian therapies may provide short-lived partial benefit for a minority of patients.

I.2. Unclarified issues related to current treatments

The majority of studies of antiparkinsonian agents have compared single agents to placebo in order to demonstrate efficacy. Few controlled studies provide evidence to guide the choice of a treatment regimen among the many existing therapies. Controlled studies have rarely compared existing therapies of any type, either those with proposed neuroprotective effects or those with symptomatic effects, either alone or in combination. In addition, little is known regarding the effects of existing therapies within clinical subgroups, defined by demographic characteristics such as age, gender or race/ethnicity or by disease features such as tremor predominance, cognitive function or age at onset. Little is known regarding the benefit of any therapeutic agent for a period of more than a few years. A review of all trials up until the end of 2001 found that the median follow-up period per trial was two years and only 40% of trials in early PD went beyond 12 months. Only two trials (DATATOP and UKPDRG) followed up patients for up to 10
years with only the latter being designed to test differences in mortality. Whether the choice of therapeutic regimen can prevent the development of adverse effects or alter the course of disease remains unresolved. Almost nothing is known regarding the relative effects of any individual therapy on survival in PD.

I.3. Needs and justification for developing new drugs

Despite the number of agents approved for use in PD, neither a risk-free treatment with sustained benefit nor a preventive or curative agent has been identified. Therefore, there is a need for new drug development in all aspects of PD therapy. Moreover, because PD and other late-life neurodegenerative disorders, such as Alzheimer’s disease, are thought to share common pathogenic features, there may be potential overlap of effect in some areas – either for primary neuroprotection, or treatment of common symptoms such as cognitive impairment or depression. Moreover, the potential number of cases of PD worldwide is expected to increase with the aging of the population. Most studies from the USA and Western Europe indicate PD prevalence to range from 1 – 2% among individuals 65 and older, and age-specific prevalence appears to double about every 5-7 years after the age of 65. The numbers of persons in the age group at risk for PD is expected to increase progressively over the next several decades in both the developed and developing world. This expected increase in the numbers of persons affected is also expected for other neurodegenerative disorders such as the atypical parkinsonisms and Alzheimer’s disease. A second, unrelated area of potential overlap for dopaminergic agents is in the treatment of restless legs syndrome, estimated to affect between 5 – 15% of the adult population.

I.4. Particularities of PD that will influence the protocol of investigation of the drug

There is no diagnostic test for PD, and diagnosis is based solely on the expertise of the examiner. The greater the expertise of the examiner, the more accurate the diagnosis, as compared to post-mortem examination, but even in the hands of experts, some diagnostic error is expected. This potential for error is greater when the PD is of short duration, and clinical signs are few. Because PD is disabling, most persons can function without therapy only within the first one or two years after diagnosis. This hampers the evaluation of agents proposed to slow, but not stop, disease progression, as the addition of symptomatic therapy confounds the evaluation of a neuroprotective effect judged by clinical criteria. Recently, imaging approaches targeting the nigrostriatal dopamine system have been proposed as adjunctive means for assessing progression of PD, although these, too, may not be independent of the potentially confounding pharmacologic effects of PD treatments. A second aspect of the requirement for symptomatic therapy in moderate or more severe PD is that the efficacy of any agent in this patient population must be evaluated as an adjunct to an established therapy, because comparison to placebo alone would not be ethical. However, comparison of a new drug to placebo is appropriate for patients receiving symptomatic therapy for the first time (sometimes called “de novo” patients), and this design is preferred by both the FDA and the EMEA for new drugs intended for registration as monotherapy. In this special setting, provision for “rescue” with a symptomatic agent may be advisable.

The response to some antiparkinsonian agents, notably those including l-dopa, can vary dramatically over time. Variation can occur over hours or even minutes during the course of a single dose, and can be further modified by conditions such as the time of day, the number of prior doses, the type of concurrent therapy, and the timing and protein content of meals. Similarly, the adverse effects of antiparkinsonian therapies, such as dyskinesias, typically wax and wane during the course of a day. These features must be taken into consideration when choosing the most appropriate outcome measure.

Newer therapies may show modest improvements in motor functions. However, without the collection of data on activities of daily living and quality of life, it is hard to evaluate whether such benefits make much difference to patients and are cost-effective.
II. PHASE II STUDIES FOR REGISTRATION OF NEW DRUGS

II.1. Outline a typical development plan

During Phase II of drug development, candidate therapies are usually tested against placebo, though occasionally other existing treatments, in the disease population of question. Generally, PD clinical trials address two primary impairments, either parkinsonism itself or motor complications that occur as part of the disease and its chronic treatment and take the form of dyskinesias (involuntary movements) or motor fluctuations (poor response to medications). For each, the primary focus can be on the delay of development of impairment or treatment of impairment once developed. Patient selection depends on the primary clinical problem being addressed. For example, studies of delay in clinical progression of parkinsonism focus on early disease and patients are usually on no other medications for Parkinson’s disease, whereas treatment protocols for dyskinesias and motor fluctuation usually focus on patients with more advanced disease who are already on multiple antiparkinsonism drugs. Almost all randomized, double-blind, controlled trials in PD are parallel in design.

II.2. Short-term studies

Short-term efficacy trials are usually three to six months with built-in titration and withdrawal phases. In some of these studies, long-term open label continuation phases are included for the acquisition of safety data.

II.2.A. Objectives

Studies usually are aimed at treating impairment and therefore focus on alleviating parkinsonism itself or improving existing dyskinesias or motor fluctuations.

II.2.B. Primary Endpoints

For parkinsonism:
   a. Comparison of the Unified Parkinson’s Disease Rating Scale (UPDRS) total score in relation to baseline scores.
   b. Comparison of the UPDRS Motor Examination (Part III) can be used as well, or the combined Activities of Daily Living and Motor Examination score (Parts II and III) in relation to baseline scores.
   c. The UPDRS is internationally utilized and has largely replaced earlier scales like the Columbia and Webster scales.
   d. The Hoehn and Yahr scale was formerly used, but it is a non-continuous scale, poorly responsive to interventions, and therefore more frequently used currently to describe patient groups and define entry criteria, rather than serving as a primary end-point.

For motor fluctuations:
   a. Dyskinesias are usually rated with the Abnormal Involuntary Movement Scale, or the Rush Dyskinesia Scale.
   b. Motor fluctuations are measured with at-home diaries for which patients undergo training in the study center on the operational definitions of “ON” (good medication response), “ON with disabling dyskinesias” (good medication response, but with superimposed involuntary movements that interfere with activities), and “OFF” (poor medication response). Reductions in overall OFF time without an increase in ON with dyskinesias indicate improved motor fluctuations. Global measures on motor fluctuations and dyskinesias can be obtained from UPDRS Part IV.

II.2.C. Secondary endpoints

For parkinsonism:
   a. Hoehn and Yahr stage, the Schwab and England rating scale
b. Dyskinesias and motor fluctuation as secondary outcomes are measured as described in Primary endpoints
c. Global scales like the Clinical Global Impression Severity and Clinical Global Impression Change scores are also used. More studies now include disease specific Quality of Life measures, such as the PDQ-39, and PDQUALIF or generic measures such as EuroQol and the SF-36. (It can be argued that quality of life measures should be primary rather than secondary outcomes)

For motor fluctuations or dyskinesias:
a. Secondary endpoints include UPDRS and all primary endpoints listed under Parkinsonism.

II.2.D. Exploratory endpoints
Depending on the chemistry of the intervention, exploratory endpoints can include non-motor elements of Parkinson’s disease such as depression, cognition, hallucinations, and dysautonomia. Rating scales for all of these are derived from standard scales in the medical literature and have not been developed specifically for Parkinson’s disease in most instances.

II.2.E. Study Design
Randomized double-blind placebo controlled parallel studies are the gold standard. After screening and entry criteria are verified, subjects are randomized and are seen regularly during study-drug intervention and then withdrawn from the drug and seen at a close-out visit. Some studies assign patients to a fixed dose (or doses) of the study drug or placebo and others allow dose titration to a maximal tolerated dose that is pre-determined. Patients are usually seen weekly for the first four weeks or through the titration phase, then at longer intervals during chronic treatment with an end-of-exposure visit and a final visit one week after drug-exposure cessation.

II.2.F. Study population
The study population depends on the question being addressed:
Parkinsonism: In early disease, the subjects are usually on no treatment for parkinsonian signs at the time of enrollment and receive monotherapy with either placebo or study drug. In moderate and advanced disease, patients are enrolled who have inadequate efficacy from their current drug therapy and can often be on levodopa or another agent with a pharmacological mechanism that is different from the study drug under question.

Motor complications: In advanced disease where the focus is usually on motor complications, patients must have dyskinesias and/or motor fluctuations of sufficient severity to warrant intervention.

Patients with another illness sufficiently severe as to jeopardize participation in the study or interpretation of results are excluded. Patients must pass screening tests for depression or dementia.

II.2.G. Specific Inclusion Criteria
Parkinson’s disease is defined clinically based on various diagnostic standards, such as the UK Brain Bank criteria. Inclusion criteria for admitting mild, moderate or advanced patients with PD are primarily based on Hoehn and Yahr stage and medication exposure. Early monotherapy studies restrict subjects to Stage I-III (unilateral or bilateral disease that may include mild balance problems but no spontaneous falling) and sometimes to Stage I-II only (no balance problems). Studies of drugs that are added to current treatment in moderate Parkinson’s disease usually restrict patients to Stage II-IV, and therefore include patients with poor balance. For studies of motor complications, inclusion criteria usually require baseline scores for the target problem sufficiently severe enough so that patients are likely to deteriorate during the trial. For dyskinesias, a minimal baseline score on the AIMS (variably 7-10) is often used, and for motor fluctuations, inclusion often requires a minimal 25% or more OFF time on diaries or the UPDRS Part IV for study entry. These criteria are introduced to avoid “floor effects”.

232
II.2.H. Specific Exclusion Criteria
In each trial, patients cannot have an allergy to the product being tested. Current exposure to various medications and past exposure to levodopa may exclude subjects from the early monotherapy trials. Because dopamine is a precursor to melanin, studies of drugs that alter levodopa bioavailability or metabolism exclude patients with a past history of melanoma. Almost no intervention studies permit Stage V patients who are wheelchair-bound at their peak treatment. Those with parkinsonian syndromes other than PD are typically excluded.

II.2.I. Tools for assessing Primary EndpointS
b. For Parkinsonism: UPDRS, total or Part III, or Parts II + III
c. For Dyskinesias: AIMS, Rush Dyskinesia Scale
d. Motor Fluctuations: Home diaries, Part IV of UPDRS

II.2.J. Specific criteria for early withdrawal and discontinuation
The investigator will discontinue a subject before completion of the study if consent is withdrawn, if the double-blind code is broken, if it is medically unacceptable to continue treatment due to adverse events or other reasons, or if pregnancy occurs.

II.2.K. Data analysis methods
The analysis of efficacy variables is based on the intention-to-treat strategy. All statistical tests are two-sided and p values <0.05 are considered statistically significant, though sometimes adjusted p-values are used to account for multiple comparisons. Many statistical methods can be used for the primary and secondary endpoints and include analysis of variance, analysis of covariance, and logistic regression.

II.3. Long-term studies
Long-term efficacy trials are usually nine months to five years.

II.3.A. Objectives
Studies are usually aimed at delaying the development of impairment, e.g. progressive parkinsonism, motor fluctuations or dyskinesias.

II.3.B. Primary Endpoints
For delay in progression of parkinsonism:
   a. Need to start dopaminergic therapy.
   b. Prespecified increase in a standard measurement tool of parkinsonism (UPDRS)

For delay in development of motor fluctuations and dyskinesias
   a. Time to development of these complications
   b. % of the population at given time points who have the complication

II.3.C. Secondary endpoints
a. For Delay in Parkinsonism Progression: secondary endpoints can include the primary endpoints for Delay in Development of Motor Fluctuations and Dyskinesias described above.
b. For Delay in Development of Motor Fluctuations and Dyskinesias, secondary endpoints can include the primary endpoints for Delay in Parkinsonism Progression described above
c. For all studies, other secondary endpoints are UPDRS scores, Hoehn and Yahr stage, Schwab and England rating scale score at specified time points.
d. For all studies, global secondary endpoints include Clinical Global Impression Severity and Clinical Global Impression Change scores as well as disease specific Quality of Life measures, such as the PDQ-39, and PDQUALIF or generic measures, such as the EuroQol and SF-36.
e. Increasingly, long-term studies are being accompanied by neuroimaging markers using beta CIT SPECT scanning or 18-F-dopa PET scanning indices.

II.3.D. Exploratory endpoints
Depending on the chemistry of the intervention, exploratory endpoints can include non-motor elements of Parkinson’s disease such as depression, cognition, hallucinations, and dysautonomia. Rating scales for all of these are derived from standard scales in the medical literature and have not been developed specifically for Parkinson’s disease in most instances.

II.3.E. Study design
Randomized double-blind placebo controlled parallel studies are the gold standard. Studies usually involve the enrollment of several hundred patients and therefore multiple centers are usually involved. After screening and entry criteria are verified, subjects are randomized and are seen usually one month after study entry and thereafter on a three or six month schedule regularly. A final visit shortly after drug-exposure cessation is standard for safety monitoring and allows the detection of withdrawal effects on primary and secondary outcomes.

II.3.F. Study population
The study population depends on the question being addressed: Delay in Parkinsonism: In early disease, the subjects are usually on no treatment for parkinsonian signs at the time of enrollment and receive monotherapy with either placebo or study drug. In moderate disease, patients are enrolled who have inadequate efficacy from their current drug therapy and can often be on levodopa or another agent with a pharmacological mechanism that is different from the study drug under question. Because these studies are long in duration, some protocols permit addition of levodopa or other drugs with continuation in the study even after the primary endpoint is reached (e.g. need for starting dopaminergic therapy) so that secondary endpoints can still be measured.

Delay in Motor complications: These studies enroll patients who at baseline are in need of dopaminergic therapy. The study randomizes patients to standard dopaminergic therapy, usually levodopa, or the new study drug. Because these studies are long in duration, some protocols permit addition of additional levodopa in both groups if inadequate efficacy of treatment is encountered in the midst of the study period.

II.3.G. Specific Inclusion Criteria
Parkinson’s disease is defined clinically with diagnostic standards, such as UK Brain Bank criteria. Other parkinsonian syndromes that are not PD are not intentionally included.

a. Delay in Progression of Parkinsonism: Inclusion criteria for long-term studies are primarily based on short duration of PD (less than five years of symptoms or diagnosis), Hoehn and Yahr stage (usually less than Stage III, meaning no significant postural reflex compromise) and medication exposure (no dopaminergic medication and often other requirements such as no Coenzyme Q, no antidepressants).

b. Delay in Development of motor complications: Inclusion criteria for long-term studies of this type enroll PD patients starting dopaminergic therapy because of need to treat the symptoms of PD. Such patients must be Stage I-III (unilateral or bilateral disease that may include mild balance problems but no spontaneous falling). Because the studies examine the onset time to motor complications, patients must not have any of these signs at baseline.

c. For both types of studies, neuroimaging data are often an intrinsic part of the protocol, so patients must be able and willing to undergo these scans, and must not have claustrophobia or other limits that preclude participation in these tests.

II.3.H. Specific Exclusion Criteria
In each trial, patients cannot have an allergy to the product being tested. Past medication exposure, especially to levodopa, may exclude subjects. Because dopamine is a precursor to melanin, some studies
exclude patients with a past history of melanoma. Those with parkinsonian syndromes other than PD are typically excluded. Patients with another illness sufficiently severe as to jeopardize participation in the study or interpretation of results are excluded. Patients who fail screening tests for dementia and depression are typically excluded.

II.3.I. Tools for assessing Primary Endpoints
a. For Delay in Parkinsonism: Clinician’s assessment of the necessity to start dopaminergic therapy (for patient safety, job security, or quality of life), UPDRS, total or Part III, or Parts II + III
b. For Dyskinesias: Onset of first dyskinesias as assessed by diaries or by Part IV of the UDPRS, AIMS, Rush Dyskinesia Scale
c. Motor Fluctuations: Onset of first OFF period by Home diaries or by Part IV of UPDRS

II.3.J. Specific criteria for early withdrawal and discontinuation
The investigator will discontinue a subject before completion of the study if consent is withdrawn, if the double-blind code is broken, if it is medically unacceptable to continue treatment due to adverse events or other reasons, or if pregnancy occurs.

II.3.K. Data analysis methods
The analyses are usually based on the intention-to-treat strategy. All statistical tests are two-sides and p values <0.05 are considered statistically significant. The primary analysis for studies involving Delay in the Development of Clinical Progression of Parkinsonism or Motor Complications evaluate survival and calculate cumulative probability of reaching each end point. Differences in outcome are expressed as hazard functions, usually with the Kaplan-Meier estimate, which can be displayed graphically. It is important to censored individuals who either drop out or have limited follow-up. Hypothesis testing between groups can use the log-rank test or Cox’s proportional hazard regression modeling, which allows adjustment for multiple covariates.

III. PHASE III STUDIES FOR REGISTRATION OF NEW DRUGS

III.1. Outline a typical development plan
Phase III trials are conducted after initial demonstration of safety and efficacy of a drug. These trials investigate larger numbers of patients, to obtain further information on efficacy and safety. A primary goal of the Phase III development plan is to obtain the information necessary for product registration. Typical studies extend the observations from Phase II studies, often employing similar study designs, but with larger numbers of subjects and longer periods of observation. Because Parkinson’s disease is relatively uncommon, Phase III studies must invariably involve multiple centers in order to allow timely accrual.

Phase III trials will address the same two primary impairments – parkinsonism, or motor complications associated with chronic treatment (dyskinesias, motor fluctuations). The aim can be either to delay or prevent the development or slow the progression of the impairment or to provide symptomatic relief.

III.2 Long-term studies to slow or halt progression of parkinsonism
Conclusive demonstration that a drug can stop or slow progression of parkinsonism remains a challenge. The design of a trial with this objective must take into account some limitations. First, there is no biomarker of Parkinson’s disease progression. Primary endpoints are based on clinical measures of parkinsonism. Because most people with Parkinson’s disease require symptomatic therapy within several months to a few years of diagnosis, such studies typically enroll only those early in the course of disease who do not require symptomatic therapy. This design avoids the confounding effect of symptomatic
treatment on the efficacy endpoint measures. No alternative marker of disease has yet been accepted as a primary endpoint to assess progression of parkinsonism.

Demonstration of efficacy in slowing or stopping progression in those with more advanced disease, who require symptomatic therapy, presents even greater challenges. Few studies have attempted to demonstrate efficacy in this population. In those with more advanced parkinsonism, clinical measures of parkinsonism are confounded by the effects of symptomatic therapy, but withdrawal of the therapy is potentially harmful. Follow-up of those with more advanced disease until death or severe disability may pose practical difficulties. Some countries are fortunate in enabling research study participants to be prospectively “flagged” on a central database (e.g. National Health Service Central Register, UK) that automatically informs researchers when a participant has died and supplies a copy of the death certificate. A typical development plan to investigate a drug proposed to slow the progression of parkinsonism will include several multicenter, double blind, placebo-controlled trials investigating the safety and efficacy of one or several doses of the drug. Monotherapy trials are the norm. Patients are recently diagnosed and not in need of symptomatic therapy. Prior exposure to symptomatic antiparkinsonian therapy is precluded, or limited to a short time, to avoid confounding. Other agents proposed to slow the progression of parkinsonism are excluded. Because this type of drug would be expected to be used for many years, or even for the entire duration of disease, safety monitoring should allow detection of events expected to occur at moderate frequency (0.5 – 5%). In general, around 1200-1500 persons exposed in short and long term studies should be adequate to detect differences of around 2 to 3% across groups. To detect adverse events occurring after prolonged use of the drug, duration of monitoring should be at least 12 months for some subjects. Depending on the specific agent, special safety monitoring may be indicated.

III.2.A. Objectives
Study goals typically are to slow or stop progression of parkinsonism

III.2.B. Primary Endpoints
Efficacy:
    a. Need to start symptomatic antiparkinsonian therapy; or
    b. A prespecified worsening in a standard clinical assessment instrument, usually the UPDRS

Safety:
    a. All adverse events;
    b. Depending on the specific agent, monitoring special safety endpoints may be indicated. For example, agents thought to block apoptotic neuronal cell death may conceivably also present an increased risk of neoplasm, and special monitoring procedures for cancer may be appropriate.

III.2.C. Secondary Endpoints
    a. Clinical assessment measures of parkinsonism, including UPDRS (if not a primary endpoint measure), Hoehn and Yahr stage, Schwab and England;
    b. Global measures such as need to start symptomatic therapy (if not a primary endpoint), Clinical Global Impression Severity or Change, and disease-specific Quality of Life measures, such as the PDQ-39 or PDQUALIF or generic measures such as EuroQol and SF-36;
    c. Neuroimaging outcomes measuring uptake of ligands specific to the dopamine system, such as $^{[123]}I\beta$-CIT (2β-carbomethoxy-3β-[4-iodophenyl]) and single photon positron emission tomography (SPECT) or $^{[18]}F$-dopa and positron emission tomography (PET) scanning indices.

III.2.D. Exploratory endpoints
    a. Endpoints targeting nonmotor features of Parkinson’s disease, such as cognition, dysautonomia, depression;
    b. Endpoints investigating response to symptomatic therapy once initiated;
c. Endpoints investigating development of the complications of chronic dopaminergic therapy (dyskinesias, fluctuations, hallucinations or psychosis)

d. Although death is a logical endpoint when investigating agents thought to alter the course of disease, the long disease duration (10 or more years on average, depending on the age at onset), makes death an impractical outcome for many drug development plans.

III.2.E. Study Design

Randomized double-blind placebo controlled parallel studies are the gold standard. Individual studies generally involve the enrollment of at least 300 patients and therefore multiple centers are needed.

Because there is inevitable subjectivity in endpoint determination, it is almost always desirable to require that the primary outcome measure be determined by the same rater, at a minimum for key time points (such as enrollment and endpoint). It may be desirable to identify specific raters within a center and/or to specify a required level of expertise with the primary efficacy measure. A blocked randomization, either by investigator or by center, is another approach to minimize the effect of between-rater variability in endpoint determination.

To avoid “unblinding” and the potential for biased endpoint assessment, two raters may be used -- a “treating” investigator who evaluates the patient at each visit, and a “blinded” investigator who determines performance on study primary outcome measures only at key visits, and is otherwise prohibited from knowledge of the subject. The use of video-assessment enables a core group of central raters to assess patients across a wide geographical distribution, but may be problematic for some impairments e.g. rigidity.

Subjects are typically assessed 1 – 6 weeks after the initiation of study drug, depending on the safety and pharmacology of the specific agent. Subsequent visits typically occur at 3-6 month intervals. Telephone follow up for safety monitoring may be planned between in-person visits. The duration of exposure for any one individual will vary depending on the proposed mechanism of the drug under development, but a minimum of 9 months follow up is though necessary to demonstrate efficacy in slowing the progression of parkinsonism.

One or more visits after drug-exposure cessation is standard for safety monitoring. When the drug has known or suspected symptomatic benefit in parkinsonism, the primary outcome measurement may be most easily interpreted only after study drug has been withdrawn. The symptomatic benefit may be mild, and only identifiable when symptoms worsen after drug withdrawal. The primary efficacy outcome may at times be determined at an interval after withdrawal of the drug under development. When planning the timing of post-treatment assessments, it will be important to consider the pharmacology of the drug under study, so that the study drug is washed out when assessments are performed.

Follow up is often continued after primary efficacy data have been obtained. Extended follow up can be especially valuable in monitoring safety, and to assess secondary outcomes such as the development of dyskinesia, motor fluctuations or psychosis.

III.2.F. Study population

Subjects typically have recently diagnosed Parkinson’s disease, not requiring symptomatic therapy and with little or no prior exposure to symptomatic or proposed neuroprotective therapies. Because Parkinson’s disease is diagnosed only by clinical criteria, and the full complement of diagnostic signs may not manifest for several years, it is expected that a percentage of those meeting diagnostic criteria with recently diagnosed Parkinson’s disease will be misclassified. Experts with greater familiarity with Parkinson’s disease have greater long-term diagnostic accuracy, but some error is inevitable. The greater potential for misclassification in early disease should be considered in determining sample size. Whilst misclassification may result in attenuated effect estimates, assuming no therapeutic benefit for the patients
who have been misdiagnosed, this result may be a more realistic estimate of treatment benefits outside trials, where expert diagnosis may not always be available prior to initiating therapy.

III.2.G. Specific Inclusion Criteria
a. Parkinson’s disease is defined clinically using published diagnostic standards, such as the UK Brain Bank and NIH criteria. These criteria exclude those with signs suggesting other parkinsonian syndromes.
b. Disease duration is typically less than five years after diagnosis. Symptom duration may be used, although this measure is dependent on patient report and may be less reliable.
c. Symptomatic treatment must not be needed in the opinion of the subject and the investigator. Enrolling subjects would need to be comfortable without symptomatic therapy if parkinsonism did not progress.
d. Mild disease severity defined using clinical criteria, such as Hoehn and Yahr stage less than III. Other clinical measures such as tremor scores or Schwab and England scores may also be used.
e. No or limited prior treatment with drugs proposed to slow disease progression (e.g., selegiline, coenzyme Q10).
f. No or limited prior treatment with symptomatic antiparkinsonian treatments.
g. If imaging is a secondary outcome, patients must be able and willing to undergo these scans, and must not have claustrophobia or other limits that preclude participation in these tests. In many cases, only a subgroup of subjects participate in the imaging study.

III.2.H. Specific Exclusion Criteria
a. Current treatment with drugs that could alleviate parkinsonism (e.g., dopaminergic, anticholinergic drugs). In early Parkinson’s disease, the appropriate washout period for the available symptomatic therapies is not well established. Generally a washout period of at least 4 weeks is desirable, in order to avoid excess early terminations due to enrollment of subjects not able to function without antiparkinsonian treatment.
b. Current treatment with drugs that could worsen (e.g., most antipsychotics, some antiemetics) parkinsonism.
c. Any serious illness that may affect participation.
d. Known allergy to study drug or related compounds.
e. Use of medication thought to interact with study drug.
f. At risk for an adverse effect of a specific drug.

III.2.I. Tools for assessing Primary Endpoints
The endpoints are determined by the investigator using clinical skills. Familiarity with the disease and the specific instruments used is therefore critical to the integrity of the study endpoint. Training in the use of the endpoint instruments is desirable.

III.2.J. Specific criteria for early withdrawal and discontinuation
The investigator will discontinue a subject before completion of the study if consent is withdrawn, if the double-blind code is broken, if it is medically unacceptable to continue treatment due to adverse events or other reasons, or if pregnancy occurs. A special case occurs when the primary endpoint measure is the investigator-determined need for symptomatic therapy. If the subject initiates symptomatic therapy prior to the investigator-determined end point, it may be preferable to continue to observe the subject on study drug and symptomatic therapy when possible, in order to obtain additional safety information.

III.2.K. Data analysis methods
The primary analyses are based on the intention-to-treat strategy. All statistical tests are two-sided. Generally p values <0.05 are considered statistically significant. The primary analysis approach typically evaluates survival and calculates cumulative probability of reaching each end point. Differences in
outcome are expressed as hazard functions, usually with the Kaplan-Meier estimate, which can be displayed graphically. It is important to censor individuals who either drop out or have limited follow-up. Hypothesis testing between groups can use the log-rank test or Cox’s proportional hazard regression modeling, which allows adjustment for multiple covariates.

### III.3 Long term studies to provide symptomatic improvement in parkinsonism

While there are a number of agents with demonstrated antiparkinsonian efficacy, none provides sustained symptomatic benefit throughout the course of this lifelong disorder. The acute and chronic side effects of established therapies are additional sources of concern. Trials of new therapies should be designed to address these concerns, with the goal of developing new drugs with more favorable efficacy and side effect profiles.

A typical development plan to investigate a drug proposed to provide symptomatic improvement of parkinsonism will include several multicenter, double blind, placebo-controlled trials investigating the safety and efficacy of one or several doses of the drug. Monotherapy trials, comparing the study drug to placebo, will in most cases be limited to early disease, enrolling “de novo” patients receiving symptomatic therapy for the first time. In advanced Parkinson’s disease, the study-drug will typically be given as adjunctive therapy, and compared to placebo given adjunctively, as it would be unethical to withdraw existing therapies. Most commonly, the efficacy of the new agent when given in combination with a dopaminergic agent (usually L-dopa plus decarboxylase inhibitor) is compared to the efficacy of placebo combined with the same agent. The addition of an adjunctive antiparkinsonian agent can result not only in improvement of parkinsonism, but also in the new onset of dopaminergic side effects, or the worsening of existing side effects, such as dyskinesias or psychosis. Specific monitoring for this possibility, and provisions for adjustment of therapies, appropriate to the specific drugs, should be included in the study protocol. In addition a global measure such as a disease specific quality of life measure is essential as it is otherwise impossible to interpret an improvement in motor function coupled with a deterioration in side effects. An alternative design compares the new drug to standard therapy. Design of such studies is often difficult due to uncertainty regarding equivalence of dosage. Some regulatory agencies may be less receptive to comparison study designs. As for all development plans, close contact with scientists in the regulatory agencies is essential.

Safety evaluations should take into account the chronic use expected for most drugs in this category. Therefore, safety monitoring should allow detection of events expected to occur at moderate frequency (0.5 – 5%). In general, around 1200-1500 persons exposed in short and long term studies should be adequate to detect differences of around 2 to 3% across groups. To detect adverse events occurring after prolonged use of the drug, duration of monitoring should be at least 12 months for some subjects. Depending on the specific agent, special safety monitoring may be indicated.

#### III.3.A. Objectives

Studies are aimed at treating impairment due to parkinsonism or improving existing dyskinesias or motor fluctuations.

#### III.3.B. Primary Endpoints

For parkinsonism:

a. Comparison of the Unified Parkinson’s Disease Rating Scale (total score) relative to baseline scores.

b. Comparison of the UPDRS Motor Examination (Part III) can be used as well, or the combined Activities of Daily Living and Motor Examination score (Parts II and III) relative to baseline scores. The UPDRS Part I includes nonmotor features and does not distinguish between primary features of disease and drug-induced side effects. For this reason, some prefer not to use Part I when assessing a new drug as adjunctive therapy along with a dopaminergic agent.

c. The UPDRS is internationally utilized and has largely replaced earlier scales like the Columbia and Webster scales.
d. The Hoehn and Yahr scale was formerly used, but it is a non-continuous scale, poorly responsive to interventions, and therefore more frequently used currently to describe patient groups and define entry criteria, rather than serving as a primary end-point.

For motor fluctuations:

a. Dyskinesias are usually rated with the Abnormal Involuntary Movement Scale (AIMS), or the Rush Dyskinesia Scale. Dyskinesias are often intermittent, and an at-home diary may be used. However, a self-report diary will likely identify only dyskinesias of moderate to severe intensity, as mild dyskinesias may be missed by the patient experiencing them.

b. Motor fluctuations are measured with at-home diaries for which patients undergo training in the study center on the operational definitions of “ON” (good medication response), “ON with disabling dyskinesias” (good medication response, but with superimposed involuntary movements that interfere with activities), and “OFF” (poor medication response). Decrease in overall OFF time without an increase in ON with dyskinesias indicate improved motor fluctuations.

c. Global measures on motor fluctuations and dyskinesias can be obtained from UPDRS Part IV.

III.3.C. Secondary endpoints

For parkinsonism:

a. Hoehn and Yahr stage, the Schwab and England rating scale

b. Dyskinesias and motor fluctuation as secondary outcomes are measured as described in Primary endpoints

c. Global scales like the Clinical Global Impression Severity and Clinical Global Impression Change scores are also used, as well as disease specific Quality of Life measures, such as the PDQ-39, and PDQUALIF and generic measures such as the EuroQol and SF-36.

For motor fluctuations or dyskinesias:

a. Secondary endpoints include UPDRS and all primary endpoints listed under Parkinsonism.

III.3.D. Exploratory endpoints

Depending on the chemistry of the intervention, exploratory endpoints can include non-motor elements of Parkinson’s disease such as depression, cognition, somnolence, hallucinations or dysautonomia. Rating scales for all of these are derived from standard scales in the medical literature. In most cases these have not been developed specifically for Parkinson’s disease.

III.3.E. Study Design

Randomized, double-blind, placebo-controlled, parallel group studies are the standard. After screening and entry criteria are verified, subjects are randomized and are seen regularly during study-drug intervention and then withdrawn from the drug and seen at a close-out visit. Some studies assign patients to one or more fixed doses or placebo. Other designs allow dose titration to an efficacy endpoint (e.g., loss of motor fluctuations) or to a pre-determined maximum, if tolerated. Visit frequency is determined in part by the pharmacologic and safety profile of the study drug, and the endpoint(s) of interest. Weekly or biweekly visits are typical during the titration phase, followed by longer between visit intervals, such as 4 – 12 weeks. There is an end-of-exposure visit and a final visit one week after drug-exposure cessation.

Because there is inevitable subjectivity in endpoint determination, it is almost always desirable to require that the primary outcome measure be determined by the same rater, at a minimum for key time points (such as enrollment and endpoint). It may be desirable to identify specific raters within a center and/or to specify a required level of expertise with the primary efficacy measure. A blocked randomization, either by investigator or by center, is another approach to minimize the effect of between-rater variability in endpoint determination.
To avoid “unblinding” and the potential for biased end point assessment, two raters may be used -- a “treat ing” investigator who evaluates the patient at each visit, and a “blinded” investigator who determines performance on study primary outcome measures only at key visits, and is otherwise prohibited from knowledge of the subject. The use of video-assessment enables a core group of central raters to assess patients across a wide geographical distribution, but may be problematic for some impairments e.g. rigidity.

Follow up is often continued after primary efficacy data have been obtained. Extended followup can be especially valuable in monitoring safety, and to assess chronic efficacy.

III.3.F. Study population
The study population depends on the question being addressed:
Parkinsonism: In early disease, the subjects are usually on no treatment for parkinsonian signs at the time of enrollment and receive monotherapy with either placebo or study drug. In moderate and advanced disease, patients are enrolled who have inadequate efficacy from their current drug therapy. Current therapy is typically levodopa or another agent with a pharmacological mechanism that is different from the study drug under question. When developing a drug for adjunctive use, the determination of what standard therapies will be acceptable must be made. Most commonly new adjunctive treatments are compared to placebo in patients receiving levodopa.

Motor complications: In advanced disease where the focus is usually on motor complications, patients must have dyskinesias and/or motor fluctuations of sufficient severity to warrant intervention.

Patients with another illness sufficiently severe as to jeopardize participation in the study or interpretation of results are excluded. Patients who fail screening tests for dementia and depression are typically excluded.

III.3.G. Specific Inclusion Criteria
Parkinson’s disease is defined clinically based on various diagnostic standards, such as UK Brain Bank criteria. Inclusion criteria for admitting mild, moderate or advanced patients with PD are primarily based on Hoehn and Yahr stage and medication exposure. Early monotherapy studies restrict subjects to Stage I-III (unilateral or bilateral disease that may include mild balance problems but no spontaneous falling) and sometimes to Stage I-II only (no balance problems). Studies of drugs that are added to standard treatment in moderate Parkinson’s disease usually restrict patients to Stage II-IV, and therefore include patients with poor balance. For studies of motor complications, inclusion criteria usually require baseline scores for the target problem sufficiently severe enough to allow change determination during the trial. For dyskinesia, a minimal baseline score on the AIMS (variably 7-10) is often used, and for motor fluctuations inclusion often requires a minimal 25% or more OFF time on diaries or the UPDRS Part IV for study entry. These criteria are introduced to avoid “floor effects”. The existing antiparkinsonian drug regimen should be optimized before determining eligibility.

III.3.H. Specific Exclusion Criteria
In each trial, patients cannot have an allergy to the product being tested. Current exposure to various medications and past exposure to levodopa may exclude subjects from the early monotherapy trials. Because dopamine is a precursor to melanin, studies of drugs that alter levodopa bioavailability or metabolism exclude patients with a past history of melanoma. Almost no intervention studies permit Stage V patients who are wheelchair-bound at their peak treatment. Those with parkinsonian syndromes other than PD are typically excluded.

III.3.I. Tools for assessing Primary Endpoints
b. For Parkinsonism: UPDRS, total or Part III, or Parts II + III
c. For Dyskinesias: AIMS, Rush Dyskinesia Scale
d. Motor Fluctuations: Home diaries, Part IV of UPDRS
III.3.J. Specific criteria for early withdrawal and discontinuation
The investigator will discontinue a subject before completion of the study if consent is withdrawn, if the
double-blind code is broken, if it is medically unacceptable to continue treatment due to adverse events or
other reasons, or if pregnancy occurs.

III.3.K. Data analysis methods
The analysis of efficacy variables is based on the intention-to-treat strategy. All statistical tests are two-sides
and p values <0.05 are considered statistically significant. Many statistical methods can be used for the
primary and secondary endpoints and include analysis of variance, analysis of covariance, and logistical
regression.

III.4. Long term studies to delay the development of dyskinesias or motor fluctuations
Dyskinesias and motor fluctuations are inevitable side effects for most patients requiring levodopa. These
side effects are much less commonly associated with other antiparkinsonian agents. However, the majority
of persons with Parkinson’s disease eventually require levodopa therapy.

III.4.A. Objectives
Studies usually are aimed at delaying the development of motor fluctuations, dyskinesias or both.

III.4.B. Primary Endpoints
   a. Time to development of these complications
   b. % of the population at given time points who have the complication

III.4.C. Secondary endpoints
   a. Measures of parkinsonian impairment, such as UPDRS scores, Hoehn and Yahr stage, Schwab and
      England rating scale
   b. Clinical Global Impression Severity and Clinical Global Impression Change
   c. Quality of Life measures, such as the PDQ-39, and PDQUALIF.
   d. Neuroimaging outcomes measuring uptake of ligands specific to the dopamine system, such as
      $^{[123]}$I-β-CIT (2β-carbomethoxy-3β-[4-iodophenyl]) and single photon positron emission tomography
      (SPECT) or $^{[18]}$F-dopa and positron emission tomography (PET) scanning indices.

III.4.D. Exploratory endpoints
Depending on the chemistry of the intervention, exploratory endpoints can include non-motor elements of
Parkinson’s disease such as depression, cognition, hallucinations, and dysautonomia. Rating scales for all of
these are derived from standard scales in the medical literature and have not been developed specifically for
Parkinson’s disease in most instances.

III.4.E. Study Design
Randomized double-blind placebo controlled parallel studies are the standard. Studies usually involve
the enrollment of several hundred patients and therefore multiple centers are usually involved. After screening
and entry criteria are verified, subjects are randomized and are seen usually one month after study entry and
thereafter at three or six month intervals. A final visit shortly after drug-exposure cessation is standard for
safety monitoring and allows the detection of withdrawal effects on primary and secondary outcomes.

III.4.F. Study population
These studies enroll patients who at baseline are in need of symptomatic antiparkinsonian therapy. The study
randomizes patients to standard dopaminergic therapy, usually levodopa, or the new study drug. Because
these studies are long in duration, some protocols permit addition of additional levodopa in both groups if
inadequate efficacy of treatment is encountered in the midst of the study period.
III.4.G. Specific Inclusion Criteria
Parkinson’s disease is defined clinically with diagnostic standards, including UK Brain Bank criteria. Other parkinsonian syndromes that are not PD are excluded. PD patients must be newly in need of symptomatic antiparkinsonian therapy, typically Hoehn and Yahr Stage II or III. Patients should have no prior exposure or very minimal prior exposure to dopaminergic drugs and should not have motor fluctuations or dyskinesias.

If neuroimaging endpoints are important to the study design, patients must be able and willing to undergo these scans, and must not have claustrophobia or other limitations that preclude participation in these tests.

III.4.H. Specific Exclusion Criteria
In each trial, patients cannot have an allergy to the product being tested. Past medication exposure, especially to levodopa, may exclude subjects. Because dopamine is a precursor to melanin, some studies exclude patients with a past history of melanoma. Parkinsonian patients who carry other diagnoses besides Parkinson’s disease are excluded.

III.4.I. Tools for assessing Primary Endpoints
- For Dyskinesias: Onset of first dyskinesias as assessed by diaries or by Part IV of the UPDRS, AIMS, Rush Dyskinesia Scale
- Motor Fluctuations: Onset of first OFF period by Home diaries or by Part IV of UPDRS

III.4.J. Specific criteria for early withdrawal and discontinuation
The investigator will discontinue a subject before completion of the study if consent is withdrawn, if the double-blind code is broken, if it is medically unacceptable to continue treatment due to adverse events or other reasons, or if pregnancy occurs.

III.4.K. Data analysis methods
The analyses are based on the intention-to-treat strategy. All statistical tests are two-sided and p values <0.05 are considered statistically significant, though sometimes adjusted p-values are used to account for multiple comparisons. The primary analysis for studies involving Delay in the Development of Motor Complications evaluates survival and calculates cumulative probability of reaching each end point. Differences in outcome are expressed as hazard functions, usually with the Kaplan-Meier estimate, which can be displayed graphically. It is important to censor individuals who either drop out or have limited follow-up. Hypothesis testing between groups can use the log-rank test or Cox’s proportional hazard regression modeling, which allows adjustment for multiple covariates.

IV. OTHER STUDIES (NEW INDICATION TRIALS, PRAGMATIC TRIALS)

IV.1 Special clinical problems
Outside of the primary motor elements of PD (parkinsonism and motor complications), PD patients experience a number of other disabilities, including hallucinations, dementia, depression, dysautonomia, sexual dysfunction, and fatigue. Drugs that are useful in for treating these symptoms in other medical contexts can be tested in PD through randomized double-blind placebo-controlled trials of PD subjects with the target problem. The example of hallucinations is provided, as a prototype, because it has been studied more than the other special clinical problems listed. For each of the other conditions, similar studies can be performed using PD patients with the target problem and appropriately designed measurement tools adapted from other medical fields.
IV.2. Hallucinations

IV.2.A. Objectives
Reduce the frequency or severity of hallucinations in drug-treated patients with chronic PD and hallucinations.

IV.2.B. Primary Endpoints
Change scores on standardized measures of hallucinations or global psychiatric disturbance.

IV.2.C. Secondary endpoints
Because drugs that improve hallucinations generally block dopamine receptors, the risk of aggravating PD is substantive and therefore secondary endpoints include standard assessments of parkinsonism, including UPDRS and Hoehn and Yahr stage.

IV.2.D. Exploratory endpoints
Scores on inventories for Depression, Cognition, and Quality of Life.

IV.2.E. Study Design
Open label exploratory and double blind placebo-controlled or clozapine-controlled trials have been conducted. These studies are usually short-term (4 weeks to 3 months), and parallel in design.

IV.2.F. Patient sample
Subjects with chronic hallucinations, with severity and frequency defined by clinical judgment as in need of treatment or by specific scores on screening tests like the Hallucination and Delusion items of the Neuropsychiatric Inventory, are enrolled in studies of agents shown to be useful against hallucinations in psychiatric populations, usually schizophrenia. Traditionally, drug dosage ranges in PD are up to 100 times less than in schizophrenia.

IV.2.G. Specific Inclusion Criteria
Parkinson’s disease is defined clinically based on various diagnostic standards, such as UK Brain Bank criteria. Inclusion criteria must establish that hallucinations began after chronic exposure to dopaminergic drugs in order to exclude the contamination of the sample by subjects with Dementia with Lewy Bodies. Entry criteria must establish that hallucinations are frequent and severe enough at baseline to warrant intervention and that scores on the baseline hallucination assessment are high enough to permit detection of change with the intervention.

IV.2.H. Specific Exclusion Criteria
In each trial, patients cannot have an allergy to the product being tested. Current exposure to other treatments for hallucinations is not permitted and usually an abstinence from such drugs for a minimum of 4 weeks is required. Almost no intervention studies permit Stage V patients who are wheelchair-bound at their peak treatment. Parkinsonian patients who carry other diagnoses besides Parkinson’s disease are excluded.

IV.2.I. Tools for assessing Primary Endpoints
Specific hallucination scales includes the Scale for Positive Symptoms (SAPS), the Parkinson Psychosis Scale, Item I (Thought Disorder) of Part I of the UPDRS, and individual items on various scales including the Neuropsychiatric Inventory. Global scales include the Brief Psychiatric Rating Scale, total Part I score of the UPDRS, and the Clinical Global Impressions scale.

IV.2.J. Specific criteria for early withdrawal and discontinuation
The investigator will discontinue a subject before completion of the study if consent is withdrawn, if the double-blind code is broken, if it is medically unacceptable to continue treatment due to adverse events or other reasons, or if pregnancy occurs.
IV.2.K. Data analysis methods

The analysis of efficacy variables is based on the intention-to-treat strategy. All statistical tests are two-sided and p values <0.05 are considered statistically significant. Multiple statistical methods can be used for the primary and secondary endpoints and include analysis of variance, analysis of covariance, and logistic regression. Clozapine is the only drug that has been shown to have efficacy for hallucinations in double-blind placebo-controlled trials but it is expensive and has potentially dangerous side effects. As it is unlikely that a new drug will be superior to clozapine but may be cheaper and/or safer, a study of a new agent against clozapine could be designed as an equivalence trial as long as it is powered to detect sufficiently narrow confidence intervals around equivalence. It is generally the case that equivalence trials require a larger sample size for the same power than a conventional trial.

IV.2. Surgical interventions: Deep brain stimulation, lesions, and cellular replacement therapies

The increased knowledge of the disrupted anatomical circuitry in PD has prompted laboratory studies and clinical trials of surgical interventions. On the premise that the degeneration in PD leaves several nuclei overactive from loss of inhibition, lesions and high voltage electrical stimulation have been applied to several deep brain structures, including the globus pallidum, thalamus and subthalamic nucleus. Alternatively, cellular replacement therapies focus on transplanting dopaminergic cells into the striatum in an attempt to reinnervate denervated structures. These studies are designed to evaluate both short-term and long-term (one to two years) efficacy, having the same objectives, rating tools and analytic methods described above for the treatment of parkinsonism and motor complications. Only special issues that are particular to these surgical trials are listed below:

IV.2.A. Study Design

Whereas randomized double-blind controlled parallel studies are the standard design in medication trials, this model is more difficult to effect in surgical trials. Lesion studies have primarily been open-label observations, and only a few have used a prospectively followed comparison group that receives optimal medical care. A few have randomized patients between two different surgical procedures. For deep brain stimulation, blinded ratings with the stimulator turned on and turned off have been used for comparisons. In cellular replacement therapies, the randomized, double-blind placebo-controlled parallel design that is typical of medical trials has been most closely replicated. In these cases, subjects are randomized between treatment groups and those who are assigned the control group go to the operating room, have a skull burr hole placed, but no needle penetration or cellular placement into the brain occurs. The surgical investigator is the only person on the research team who is unblinded, and all ratings are performed by investigators who were not involved in the surgery. In all studies, subjects are evaluated at baseline (often with more than one baseline assessment) and then seen regularly after surgery, usually at one month, and every three months thereafter during the trial. The score at the final visit usually serves as the primary outcome measure having adjusted for baseline scores. Often the percentage change in baseline scores is presented across different interventions. A common feature of all these trials is that the sample size is much smaller given the complexity and expense of the interventions. This makes the use of standardized outcome measures even more important as, inevitably, pooling results through the use of meta-analysis will be required to reduce the likelihood of both type I and type II errors. One special feature of surgical trials is the ability to assess an intervention undertaken either unilaterally or bilaterally. If subjects are randomised to have a different procedure for each side then, they can act as their own controls and matched methods of analysis are required as in cross-over studies. More typically comparisons are made by side so that 20 patients treated bilaterally provide 40 outcome measures. In this case, it is important to remember that each observation is not truly independent as they clustered within individuals and more complex statistical methods are required to allow for this clustering.
IV.2.B. Study population
The study population for surgical interventions are subjects with advanced PD who have failed other therapies, but still show an objective improvement (even if for short intervals) to dopaminergic therapy. They have motor complications in the form of dyskinesias and/or motor fluctuations. During “on” periods, they must be Hoehn and Yahr Stage I-III and during “off” periods, they must be Stage III-V.

IV.2.C. Specific Inclusion Criteria
Because surgical intervention is a major medical treatment, patients must be in good health other than their PD. Most studies require good cognitive status (MMSE usually at least 24) and no hallucinations. They must clearly understand all surgical risks and have a caregiver who can participate in the program.

IV.2.D. Specific Exclusion Criteria
Parkinsonian patients who carry other diagnoses besides Parkinson’s disease are excluded. Dementia, hallucinations and other significant behavioral problems usually exclude patients from these trials.

V. SPECIAL CONSIDERATIONS
Placebo effects are frequent and substantive in PD trials. The dopamine system is directly involved in the regulation of reward mechanisms, expectation, motivation and vigilance. Positron emission tomography [\(^{11}\text{C}\)]-raclopride binding studies document evidence of increased striatal dopamine release in PD subjects responding to placebo treatment. Because most drugs or interventions being studied in PD share dopaminergic augmentation mechanisms, separating primary dopaminergic effects due to the intervention vs. dopaminergic effects due to study participation (placebo effects) must be delineated. Whereas a positive effect on parkinsonism is anticipated in both the control and study group in PD, the outcome scores, after adjusting for baseline scores, must be significantly better in the intervention arm than in the placebo-treated arm before one can conclude that improvement is due to the intervention.

VI. EXAMPLES OF LANDMARK TRIALS

Treatment of parkinsonism

Treatment of motor complications

Prevention of clinical progression of parkinsonism

Prevention of disease mortality
Prevention of development of motor complications


Treatment of special issues


Surgical interventions


VII. SUGGESTED READINGS

3. Tanner CM, PD or not PD?—that is the question. Neurology 2003;61:5-6.
Chapter 20. Multiple Sclerosis and Other Demyelinating Diseases

Giancarlo Comi, M.D. ¹
Michel Clanet, M.D. ²
Fred Lublin, M.D. ³
Vittorio Martinelli, M.D. ⁴
Chris H. Polman, M.D. ⁵
Per Solberg Sorensen, M.D. ²

¹ Department of Neurology
University Vita-Salute IRCCS San Raffaele
Via Olgettina, 60
20132 Milan
ITALY
Tel  +39/02/26432990
Fax  +39/02/26433746
Email  comi.giancarlo@hsr.it

² Department of Neurology
Hospital Purpan
Toulouse
FRANCE

³ Department of Neurology
VU Medical Centre
Amsterdam
THE NETHERLANDS

⁴ Department of Neurology
University Vita-Salute IRCCS
San Raffaele, Milan
ITALY

⁵ Department of Neurology
Copenhagen University Hospital
DENMARK

Corresponding Author:
Giancarlo Comi, M.D.
I. INTRODUCTORY REMARKS

Multiple Sclerosis (MS) is a devastating disease affecting predominantly the white matter of the central nervous system. It is the most frequent cause of disability in young adults, after car accidents. The disease affects about one million people, mostly in developed countries. In fact the disease has a peculiar geographic distribution, probably as a result of an interaction of genetic and environmental factors, being more frequent in North America and North Europe and almost absent in equatorial regions and in Asia. Environmental factors are completely unknown and genetic factors have been only partially revealed, with a key role for genes involved in regulatory mechanisms of the immune system. No preventing strategies are available.

In 1993 FDI approved the use of Interferon Beta 1b for the treatment of relapsing remitting MS. Shortly later both Interferon beta 1a and Interferon beta 1b were approved by drug agencies worldwide and Glatiramer Acetate was approved in the end of the nineties. Finally Interferon beta 1b was approved also for treatment of secondary progressive multiple sclerosis. Soon after the approval of these immunomodulating agents, new phase II-III clinical trials were started to explore the best dose and frequency of injection for available therapies, to evaluate combination treatments and to tests safety and efficacy of new treatments. Immunosuppressive agents have been extensively used to treat MS since more than thirty years, however only recently mitoxantrone has been approved by FDI for treatment of active MS patients not responding to immunomodulating agents.

There are converging and convincing evidences that early treatment is more effective in MS and that both Interferons (IFNs) and Glatiramer Acetate (GA) produce little or no benefits in the progressive phases of the disease. Multi-weekly injections, particularly in the initial period of treatment, resulted superior to the weekly injection. However increasing dose and frequency of injection result in a higher frequency of anti-interferon antibodies, which, if persistently present may limit the efficacy of treatment.

New therapies have a broad range of targets, including the T cell receptor, the co-stimulation molecules, the blood-brain barrier permeability, chemokines, etc. It has been estimated that about one hundred clinical trials are ongoing in Europe and North America. Most of these trials have a placebo control group, which raises ethical and practical issues because of the availability of active treatments. Ethical concerns for placebo arm are particularly relevant for phase III clinical trials because the study duration is usually not shorter than 2 years.

Combination therapy, an approach derived by treatment of tumours, is already in use in some MS centres, however the efficacy of this therapeutic strategy has not been proved yet. Two approaches are generally used: 1) the comparison between treatment A and the combined treatment A+B (where treatment A is Glatiramer Acetate or Interferon beta 1b); 2) treatment A is compared to treatment B followed by treatment A (where treatment B is usually an immunosuppressive agent).

MS is not a homogenous disease. Three main courses have been described, relapsing-remitting (RR), secondary progressive and primary progressive, to which we should add atypical variants such as Marburg’s disease, Devic’s disease and Balo’s disease. Post mortem and biopsy studies revealed that MS subtypes are characterised by different pathological findings and that in the same patient pathogenetic mechanisms may vary along the disease course. This intra- and inter-individual variability of the dysfunction of the immune system may be one of the reasons why the response to immunomodulators is so variable in MS patients. It is possible that future trials will concentrate on a more homogenous group of patients using a combination of clinical, instrumental and genetic parameters.

The efficacy of IFNs and GA on MS disease activity, at least in patients with RR disease is definitely proved. However, also in this population of patients, it is still debated if the “anti-inflammatory” activity of IFNs and GA also produce long term effects on disability. There are some contradictions among
clinical trials performed in RRMS on the effects on disability, explained by methodological problems (problems in the definition of progression, sample size, study duration, poor responsiveness of EDSS etc.) and by differences in the included population. Long term placebo controlled studies in RRMS, so necessary to prove trial benefits for the prevention of the accumulation of irreversible nervous damage, are not possible anymore for ethical and practical reasons. On the other hand, comparison with epidemiological data are very difficult to interpret. The open or blind extension of placebo controlled clinical trials with the comparison of patients with early and delayed active treatment may give problems of interpretation of the results because of the patients lost to the follow up and the frequency of patients switched to other treatments.

In Secondary Progressive MS all clinical trials failed to show efficacy and only the European Multicenter trial was able to demonstrate a significant effect of IFN beta1b on the proportion of patients with confirmed increase of EDSS score. No significant effects on disability were observed in clinical trials evaluating the effects of IFNs and GA in Primary Progressive MS.

The problem of paraclinical endpoints has been widely discussed in many meetings. Their use for phase II clinical trials in RR patients, has been accepted by national and supranational drug agencies because active MRI lesions are a more precise and sensitive measure of disease activity than the count of the number of relapses. In phase III trials in RRMS and in progressive MS courses only disability measures are accepted as primary endpoint; MRI surrogate markers may be used only as secondary or exploratory measures.

**II. PHASE II STUDIES FOR REGISTRATION OF NEW DRUGS**

Phase II studies in MS have been mostly short term, i.e., approximately six months and designed to answer two main questions: 1) proof of principle of the proposed agent and 2) dosing information. Safety monitoring, as always, is also a prime consideration. To date, most phase II MS clinical trials have involved immunomodulatory or immunosuppressive agents. We should soon see trials of neuroprotective agents. For the immunomodulating agents, the basis of the trials often comes from either animal models, such as experimental allergic encephalomyelitis, or from other therapeutic areas where there is a proposed autoimmune pathogenesis, such as rheumatoid arthritis or Crohn’s disease. This strategy centers on altering the immune response to a more suppressive bias, i.e., from Th1 to Th2 or from helper cytokines to suppressive cytokines. Such an approach has advantages over a more generalized immunosuppressive approach, but there are ongoing and planned immunosuppressive strategies, as well. Recent results suggest that such translation is not without some hazard and the need to assure that an agent does not worsen MS is paramount. The problem lies, in part, with our limited understanding of the full mechanisms underlying the production of flare ups and deficit production in MS, such that therapeutic approaches considered potentially beneficial have turned out to produce the opposite result, e.g., anti-TNF directed therapies. Therefore, safety assessments must not only look for adverse events from an experimental agent but also assess whether the agent might have a negative effect on disease course. This can be monitored by either the relapse rate (looking for an increase), change in level of disability (a more difficult outcome to monitor in phase II trials) or increase in MRI activity. The latter is the easiest to accomplish, for the reasons detailed below, and has been utilized as a primary outcome measure in safety trials of combination therapy in MS. The logical basis for this is that most current MS therapies produce a reduction in the number of gadolinium enhancing lesions. Therefore, if the addition of another immunomodulating agent produced an increase in the number of gadolinium enhancing lesions would indicate an adverse interaction between the two agents.

For relapsing-remitting MS, phase II trials most commonly measure a relative reduction in gadolinium enhancing lesion activity as the indication of efficacy. As gadolinium enhancing lesions occur 6-10x as frequently as clinical exacerbations, they provide greater power for detecting differences. This power,
through frequent scanning, provides for an ease of detection and a need for smaller numbers of patients than a clinical outcome would require. In such trials, relapse rate reduction is an important secondary outcome. As these trials usually are of six months duration, changes in EDSS are less likely to be seen. Use of gadolinium enhanced lesions is only reasonable when the putative mechanism of action of the tested agent is expected to impact on the blood-brain barrier (BBB). Such would be the case with most anti-inflammatory agents and adhesion molecule blockers. Agents that might act away from, or independent of the BBB or that might not affect inflammation should have outcome measures that either reflect a clinical change, such as relapse rate, or use MRI metrics that relate to tissue damage.

In primary progressive (PP) or secondary progressive (SP) MS, phase II trials should also aim for a clinical outcome or change in an MRI metric of tissue damage, such as atrophy, T1 black hole volume, NAA spectroscopy or magnetization transfer imaging. Except for early SP MS where there may be concomitant frequent relapses, gadolinium enhancement would not be useful, nor would measures of relapse rate. One difficulty in trying to alter progressive disease lies in the clinical measurement of progression of disability. The commonest scale, the Kurtzke expanded disability status scale (EDSS) is rather insensitive, requires a minimum of six months to demonstrate a change, is weighted heavily toward ambulation and is not linear. A newer scale, the Multiple Sclerosis Functional Composite (MSFC) has several advantages over the EDSS, including linearity, need for fewer subjects to achieve the same power, and less variability, but, as opposed to the EDSS, has not yet been accepted as an outcome measure by regulatory authorities.

II.1 Outline of a typical development plan

A typical phase II trial testing an anti-inflammatory agent in RR MS will utilize a multicenter, randomised, double blind, placebo-controlled, parallel group design.

II.2. Short term studies

II.2.A. Objectives
To evaluate short term efficacy and safety of immunomodulatory or anti-inflammatory drug.

II.2.B. Primary Endpoints
Cumulative total number of enhancing lesions on all post Gd T1 weighted MRI images from monthly scans performed from week 0-12 to week 24-36. Other endpoints related to MRI disease activity can be selected as primary endpoints (see secondary endpoints).

II.2.C. Secondary Endpoints
MRI parameters:

a. number of new enhancing lesions
b. total volume of enhancing lesions
c. number of new T2 weighted lesions
d. number of T2 weighted lesions
e. total volume of T2 weighted lesions
f. number and volume of T1 weighted hypointense lesions
g. other MRI parameters such as progression of brain atrophy, or variation of magnetisation transfer ratio parameters, (these last parameters particularly useful when testing a non anti-inflammatory drug).

Clinical parameters:

a. Related to clinical relapses: number of relapses, relapse rate, time to first relapse, number of relapse-free patients
b. Related to disability progression: number of patients with a predefined interval EDSS progression (usually EDSS score ≥ 1), variations of MSFC scores.

Safety and tolerability parameters: incidence and prevalence.

II.2.D. Exploratory endpoints
Exploratory endpoints must be added according to the pharmacological properties of the drug.

II.2.E. Study design
Multicenter, randomised, double blind, placebo-controlled, parallel group design, is used, the number of arms depending on the different dosages which are evaluated. A screening phase is generally required, 4-12 weeks, with 1 or 2 MRI examinations for selecting MRI active patients. Enrichment of the population study with only MRI active patients enhances the power of the study and reduces the number of patients required. The duration of the study varies between 24 to 36 weeks, depending on the onset of monthly MRI evaluation. In some trials patients are evaluated monthly from baseline, but sometimes, when the full effect of the drug is delayed, patients are evaluated from 12th week to 36th week. A longer period of evaluation is not possible with a placebo arm. Such a design becomes more difficult to perform than a few years ago for ethical reasons and availability of MS patients.

Medication permitted is corticosteroid treatment for MS relapses: 1g methylprednisolone over 3 hours IV infusion/day for 3 or 5 days is usually the recommended treatment of severe relapses. All symptomatic treatments such as antispastic, anticholinergic, antidepressant, antiepileptic drugs and rehabilitation are usually permitted.

II.2.F. Study Population
Sufficient data are available to calculate accurately the number of patients required for the study according to the anticipated drug effect. As an example, 60 patients per group would be sufficient over 6 months to detect a 50% difference with power taken as 90% and a type one error =0.05 based on the hypothesis of 2.8 ± 3.7 new Gd enhancing T1 weighted lesions (see Sormani et al.).

II.2.G. Specific Inclusion Criteria
   a. Patients who meet the diagnosis criteria for MS according to guidelines provided by McDonald et al.
   b. Patients who present at least one T1 weighted Gd enhancing lesion on MRI performed in the screening period.
   c. Patients with clinical disability measured by EDSS score between 0 and 5.5 inclusive.
   d. Male and female MS patients aged between 18 – 55 years. Women must use an effective method of contraception if they are childbearing potential.

II.2.H. Specific Exclusion Criteria
   a. Patients who present a progressive evolution course defined as a sustained progression of disability evaluated by EDSS score in the year preceding the screening period.
   b. Patients with relapse in the 2 months period preceding baseline. This period may vary according to the drug.
   c. Patients with previous treatment with immunosuppressive, immunomodulating or any investigational drug. According to this previous treatment and the drug tested, exclusion must be complete or a wash-out period of variable duration may be accepted.
   d. Usual exclusion criteria such as patients with concomitant severe or unstable non neurological disease which would induce any risk for the patient.
   e. Pregnant or breastfeeding women.
II.2.I. Specific criteria for early withdrawal and discontinuation
A list of withdrawal criteria is pre-established such as the following: consent withdrawn, severe progression of the disease requiring recommended treatments, serious adverse event other than a relapse, insufficient compliance to the treatment, inadequate concomitant therapy, occurrence of pregnancy.

II.2.J. Data analysis methods
For efficacy analysis the primary population is the intention to treat (ITT) population. The secondary populations is the per protocol population with no major protocol deviation. For safety analysis the exposed population is analysed. Multiple statistical methods are usually used for the primary and secondary end points. The tests are usually two-sided with a global type 1 error, $\alpha \leq 0.05$. These methods, adapted to the parameters analysed, must be predefined when the protocol is designed.

II.2.K. Extension studies
The goal of phase 2 studies are: confirmation of “a proof of concept” and/or safety requirements. In the majority of cases, these controlled trials are randomized against placebo, which forbids a long term study. There is no guideline to recommend for managing the patients at the end of the treatment period, the positions of ethical committees varying from country to country.

II.3. Phase 2 studies in primary progressive MS
PPMS remains the only subtype of MS for which there is no approved disease modifying therapy. Immunomodulators and immunossuppressive therapies are used in relapsing secondary progressive MS but these compounds are not effective in SP with sustained progression without relapses, which is similar to PPMS.

Tolerability and efficacy phase 2 studies are randomized, double blind, placebo controlled trials. MRI outcomes are not accepted as surrogate outcomes for phase 2 studies in PPMS Therefore the primary outcome generally used is the time to sustained treatment failure as defined by progression on disability scale, usually EDSS or a more sensitive scale, MSFC.

Some secondary MRI outcomes are available such as the measurement of progression of brain atrophy. The duration of the treatment period is 2 years. Methodology is as the whole similar to the phase 2 trials in RRMS.

III. PHASE III STUDIES FOR REGISTRATION OF NEW DISEASE MODIFYING DRUGS

III.1. Long-term studies to slow or halt relapsing-remitting MS

III.2. Outline of a typical developmental plan
A clinical developmental plan will include at least one large phase III study with a clinical primary efficacy outcome, either relapse rate or disease progression, with a study duration of at least 2 years. The pivotal trial for registration purposes has hitherto included one large well-conducted placebo-controlled trial. Initially, both the U.S. Food and Drug Administration (FDA) and the European Medicines Evaluation Agency (EMEA) have granted approval after one adequate and well controlled trial in patients with relapsing-remitting MS. However, for the latest approved disease-modifying drug in the United States, approval was granted after one large placebo-controlled trial and a short-term comparative trial with the brand leader. In relapsing-remitting multiple sclerosis placebo-controlled trials have become increasingly difficult to perform after approval of several disease modifying drugs for treatment of the disease activity. There are both ethical and practical issues involved. The ethical problems regarding
placebo-controlled trials in relapsing-remitting multiple sclerosis are in principal identical for phase II and phase III trials, whereas the practical problems with enrolment of large patient numbers for prolonged studies are by far more pronounced in phase III studies. Regarding ethical issues for future placebo-controlled clinical trials an international taskforce of clinicians, statisticians, ethicist and regulators concluded that placebo-controlled clinical trials in forms of multiple sclerosis for which partially effective therapies exist were ethical as long as study subjects were fully appraised of the availability of such therapies and were encouraged to pursue them outside of a clinical trial. Patients who declined to utilise available treatment after proper education and counselling, or those that failed all therapies can be considered to have no treatment alternatives and thus may participate in a placebo-controlled trial. Future requirements for approval of new disease modifying drugs in relapsing-remitting MS by the FDA or EMEA are not known. A developmental plan may include both superiority head to head trials and non-inferiority trials against an established approved drug. Regulatory authorities have not yet approved placebo-controlled trials with historical placebo-controls.

III.2.A. Objectives
To evaluate the efficacy and safety of the investigational drug as mono-therapy in patients with relapsing-remitting multiple sclerosis.

III.2.B. Primary Endpoints
The primary endpoint should be clinical. Time to progression in disability should be preferred in trials of 2-3 years duration. Progression is usually measured as increase of one full (1.0) step on Kurtzkes EDSS scale (0.5 step in patients with a baseline EDSS of 5.5 or above). Progression should be confirmed at 2 assessments with an interval of 3 or 6 months. Alternatively the multiple sclerosis functional composite (MSFC) be used, but this so far has no been accepted as a valid primary endpoint by regulatory authorities, mainly because it has not yet been defined how progression is measured on this scale, and how worsening on this scale should be explained clinically. Changes in the annual relapse rate can also be used as primary endpoint. Registration of confirmed relapses is preferred to the use of reported relapses with or without confirmation.

III.2.C. Secondary Endpoints
If progression is used as the primary endpoint, the annual relapse rate should be included as secondary endpoint, and vice versa. A surrogate marker, MRI, should always be included as secondary endpoint. Gadolium-enhancing lesions on T1-weighted images, new disease activity on T2-weighted images or combined unique activity (CUA), i.e. a combination of new or enlarging T2-lesions and gadolinium positive lesions, are recommended. Alternatively, changes in T2-lesion area or T1-lesion area (black holes) can be used. Recently, a brain atrophy measure e.g. the brain parenchymal fraction has gained use as a MRI secondary endpoint.

III.2.D. Exploratory Endpoints
Clinical exploratory endpoints are time to first relapse, proportion of relapse-free patients, integrated disability status score (IDSS), i.e. the area under the disability time curve, proportion of patients with progression, and time to EDSS 6 or 7. Exploratory MRI endpoints include some of the MRI endpoints measured under secondary endpoints, gadolinium enhancing lesions, and new magnetic resonance techniques like magnetisation transfer ratios and magnetic resonance spectroscopy.

III.2.E. Study Design
A randomised double-blind, placebo-controlled parallel group design is used. The trial involves a baseline evaluation with control for fulfilling of inclusion and exclusion criteria. Patients should be assessed clinically with intervals of 3 months and in case of an acute relapse. Assessment includes scoring on the expanded disability status scale (EDSS) and/or the multiple sclerosis functional composite (MSFC) scale. MRI measures should be obtained at yearly intervals or at least at baseline and
study end. The duration of the study should be at least 2 years and preferably 3 years, if disease progression is the primary endpoint.

III.2.F. Planned sample
The sample size depends on the chosen minimal relevant deficit in the primary endpoint. Typically a sample size of above 300 patients per treatment-arm is required to detect 30-40% difference between the trial-arms with an 80% power and a type 1 error (two-sided) of 5%.

III.2.G. Study population
Patients with relapsing-remitting multiple sclerosis according to accepted criteria (McDonald criteria or Poser criteria) and age 18-55 years. Only patients with low or moderate disability (EDSS ≤ 5) should be included. Enrolled patients should have suffered recent disease activity, usually in the previous year, either clinical activity or MRI activity.

III.2.H. Specific inclusion criteria
Similar to those described for phase II placebo-controlled trial.

III.2.I. Specific exclusion criteria
Similar to those described for phase II placebo-controlled trial.

III.2.J. Specific criteria for early withdrawal and discontinuation
Placebo-controlled trials should include exit (escape) criteria defined as significance disease progression, e.g. 2 steps on EDSS, or frequent and severe relapse activity.

III.2.K. Data analysis methods
The analysis of the primary efficacy variable is based on the intention-to-treat (ITT) population. Time to progression is analysed using the log-rank test and Kaplan-Meier estimates.

IV. COMPARATIVE STUDIES

IV.1. Superiority studies
Criteria for conducting comparative superiority trials in relapsing-remitting multiple sclerosis have not been defined. There is currently no international gold standard for such trials. Below is an example that was accepted by the U.S. Food and Drug Administration as additional trial to a placebo-controlled study for approval of a disease-modifying agent. It has to be recognised that this study has been criticised in the medical and scientific community, mainly because its length was thought to be suboptimal.

IV.1.A. Objectives
To evaluate the comparative efficacy and tolerability of the investigational drug versus an approved active control in relapsing-remitting multiple sclerosis.

IV.1.B. Primary endpoints
Proportion of patients who remained relapse-free at 24 weeks.

IV.1.C. Secondary endpoints
Mean number of relapses per patient during 24 weeks, number of active lesions per patient per scan at 24 weeks on MRI.
IV.1.D. Exploratory endpoints
Mean number of combined unique activity lesions, i.e. a combination of new or enlarging T2-lesions and gadolinium positive lesions, per patient per scan, mean number of T1 active lesions per patient per scan, mean number of T2 active lesions per patient per scan, proportion of relapse-free patients at 48 weeks.

IV.1.E. Study design
A randomised parallel-group, single-blind study. Patients and treating physicians were aware of treatment allocation, whereas the evaluating neurologist and radiologist were blinded to study treatment. Ideally, the study should be double-blind but this might be difficult to achieve depending on the characteristics of the drugs under study. The primary efficacy endpoint was assessed at 24 weeks but the study was continued for 48 weeks.

IV.1.F. Planned sample
A sample size of 280 patients per treatment arm provided a 95% power at a significance level of 5% to detect a 30% relative increase in the primary endpoint.

IV.1.G. Study population
Patients with relapsing-remitting multiple sclerosis according to Poser criteria and age 18-55 years. Only patients with EDSS 0 to 5.5 and 2 relapses in the prior 2 years were included.

IV.1.H. Data analysis method
The analysis of the primary efficacy variable was based on the intention-to-treat (ITT) population. The primary end point, the odds ratio for remaining relapse free at 24 weeks, was analyzed by logistic regression with adjustment for treatment and centre.

IV.2. Non-inferiority studies

IV.2.A. Objectives
To evaluate the efficacy and tolerability of the investigational drug in patients with relapsing-remitting multiple sclerosis in comparison with an established drug at fully effective dosage under mono-therapy conditions.

IV.2.B. Primary endpoints
Primary endpoints for such studies have not been defined but should be a clinical measure assessed in the per-protocol (PP) population. Possible primary endpoints would include time to first relapse, proportion of relapse-free patients, annual relapse rate or time to progression on expanded disability status scale (EDSS).

IV.2.C. Secondary endpoints
Clinical endpoints include: Time to first relapse, relapse-free patients, annual relapse rate, time to progression on EDSS confirmed at 6 months, proportion of patients with progression.

MRI endpoints include: New disease activity on T2-weighted images or combined unique activity (CUA), e.g. a combination of new or enlarging T2-lesions and gadolinium positive lesions. Alternatively, changes in T2-lesion area, T1-lesion area (black holes), or brain atrophy can be used.

IV.2.D. Study design
The trial may be a multi-centre, double-blind, randomised parallel-group design with a double dummy technique comparing the investigational drug with the best reference treatment at optimised dosage. The double-blind phase may be followed by an open-labelled extension study.
IV.2.E. Planned sample
The authors are not aware of widely accepted sample size calculations for this type of study; it has to be recognised, however, that by concept non-inferiority studies involve a huge number of patients.

IV.2.F. Study population
Patients with relapsing-remitting multiple sclerosis according to accepted criteria (McDonald criteria or Poser criteria) and age 18-55 years. Only patients with low or moderate disability (EDSS ≤ 5) should be included. Enrolled patients should have suffered recent disease activity, usually in the previous year, either clinical activity or MRI activity.

IV.2.G. Specific criteria for early withdrawal or discontinuation
Placebo-controlled trials should include exit (escape) criteria defined as significance disease progression, e.g. 2 steps on EDSS, or frequent and severe relapse activity.

IV.2.H. Data analysis method
In non-inferiority trials, analysis of the primary efficacy variable is made on the per-protocol (PP) population. Relapse-free rates or progression-free rates may be compared by a logistic regression model whose 95% confidence interval computation may include baseline characteristics as factors.

V. SECONDARY PROGRESSIVE MS

V.1. Outline of a typical developmental plan
The benefit of disease modifying therapies in patients with secondary progressive MS is less apparent. The results in the European study with interferon-beta 1b showed a modest slowing of progression in secondary progressive MS patients of whom many had relapses. By contrast, no effect on disability progression was observed in the North American study of interferon-beta 1b or in the SPECTRIMS study with interferon-beta 1a. In another study with interferon-beta 1a, the IMPACT study, only an effect on the MSFC was found. Hence, it can be concluded that placebo-controlled double-blind trial are still ethical and feasible in patients with secondary progressive multiple sclerosis. Patients with secondary progressive MS who have still relapses should be informed about the possibility of starting with an approved drug outside a clinical trial, and only patients who have declined to do so should be included in clinical trials.

V.2. Placebo-controlled trials

V.2.A. Objectives
To evaluate the efficacy and safety of the investigational drug as mono-therapy in patients with secondary progressive multiple sclerosis.

V.2.B. Primary endpoints
The primary endpoint should be clinical. Time to progression in disability should be preferred in trials of 2-3 years duration. Progression is usually measured as increase of one full step on Kurtzkes EDSS scale (0.5 step in patients with a baseline EDSS of 5.5 or above). Progression should be confirmed at 2 assessments with an interval of 3 or 6 months. In the future the multiple sclerosis functional composite (MSFC) might provide a useful alternative, but so far this scale has not been accepted as primary outcome measure by regulatory authorities.

V.2.C. Secondary endpoints
The annual relapse rate should be included as secondary endpoint. A surrogate marker, MRI, should always be included as secondary endpoint. New disease activity on T2-weighted images or combined unique activity (CUA), e.g. a combination of new or enlarging T2-lesions and gadolinium positive lesions
or brain atrophy e.g. measured as the brain parenchymal fraction, are recommended. Alternatively, changes in T2-lesion area or T1-lesion area (black holes) can be used.

V.2.D. Exploratory endpoints
Clinical exploratory endpoints are proportion of patients with progression, and time to EDSS 6 or 7, integrated disability status score (IDSS), e.g. the area under the disability time curve, time to first relapse, proportion of relapse-free patients. Exploratory MRI endpoints may include some of the MRI endpoints measured under secondary endpoints, gadolinium enhancing lesions, and new magnetic resonance techniques like magnetisation transfer ratios and magnetic resonance spectroscopy.

V.2.E. Study design
A randomised double-blind, placebo-controlled parallel group design is used. The trial involves a baseline evaluation with control for fulfilling of inclusion and exclusion criteria. Patients should be assessed clinically with intervals of 3 months and in case of an acute relapse. Assessment includes scoring on the expanded disability status scale (EDSS) and/or the multiple sclerosis functional composite (MSFC) scale. MRI measures should be obtained at yearly intervals or at least at baseline and study end. The duration of the study should be at least 2 years and preferably 3 years, if disease progression is the primary endpoint.

V.2.F. Planned sample
The sample size depends on the chosen minimal relevant deficit in the primary endpoint. Typically a sample size of above 300 patients per treatment-arm is required to detect 30-40% difference between the trial-arms with an 80% power and a type 1 error (two-sided) of 5%.

V.2.G. Study population
Patients with secondary progressive multiple sclerosis according to accepted criteria and age 18-55 years. Only patients with moderate disability (EDSS 3 to 5.5) should be included. Enrolled patients should have suffered recent clinical disease activity, i.e. progression or relapses during the last 1-2 years.

V.2.H. Specific inclusion criteria
Similar to those described for phase II placebo-controlled trial.

V.2.I. Specific exclusion criteria
Similar to those described for phase II placebo-controlled trial.

V.2.J. Specific criteria for early withdrawal and discontinuation
Placebo-controlled trials should include exit (escape) criteria defined as significance disease progression, e.g. 2 steps on EDSS.

V.2.K. Data analysis methods
The analysis of the primary efficacy variable is based on the intention-to-treat (ITT) population. Time to progression is analysed using the log-rank test and Kaplan-Meier estimates.

VI. PRIMARY PROGRESSIVE MS

VI.1. Outline of a typical developmental plan
The clinical course in primary progressive MS is characterized by a progressive accumulation of neurological deficits from onset without relapses. There is no approved therapy for this course of multiple sclerosis.

A clinical developmental plan will include at least one large phase III study with disease progression as the clinical primary efficacy outcome, and with a study duration of at least 2 years.
VI.2. Placebo-controlled trials

VI.2.A. Objectives
To evaluate the efficacy and safety of the investigational drug as mono-therapy in patients with secondary progressive multiple sclerosis.

VI.2.B. Primary endpoints
The primary endpoint should be clinical, and time to progression in disability should be preferred in trials of 2-3 years duration. Progression is usually measured as increase of one full step on Kurtzkes EDSS scale (0.5 step in patients with a baseline EDSS of 5.5 or above). Progression should be confirmed at 2 assessments with an interval of 3 or 6 months. Alternatively the multiple sclerosis functional composite (MSFC) may be used, but it has not yet been approved as primary outcome measure.

VI.2.C. Secondary endpoints
A surrogate marker, MRI, should always be included as secondary endpoint. The currently recommended measures for therapeutic trials in relapsing remitting and secondary progressive multiple sclerosis show only little change in primary progressive multiple sclerosis and therefore more pathologically specific MRI measures may be required. Changes in T1-lesion area (black holes), brain atrophy, e.g. measured as the brain parenchymal fraction or cervical cord cross-sectional area may be used.

VI.2.D. Exploratory endpoints
Clinical exploratory endpoints are proportion of patients with progression, and time to EDSS 6 or 7. Exploratory MRI endpoints may include some of the new magnetic resonance techniques like magnetisation transfer ratios and magnetic resonance spectroscopy.

VI.2.E. Study design
A randomised double-blind, placebo-controlled parallel group design is used. The trial involves a baseline evaluation with control for fulfilling of inclusion and exclusion criteria. Patients should be assessed clinically with intervals of 3 months and in case of an acute relapse. Assessment includes scoring on the expanded disability status scale (EDSS) and/or the multiple sclerosis functional composite (MSFC) scale. MRI measures should be obtained at yearly intervals or at least at baseline and study end. The duration of the study should be at least 2 years and preferably 3 years, if disease progression is the primary endpoint.

VI.2.F. Planned sample
The sample size depends on the chosen minimal relevant deficit in the primary endpoint. It has been reported that in secondary progressive multiple sclerosis, a sample size of above 300 patients per treatment-arm is required to detect 30-40% difference between the trial-arms with an 80% power and a type 1 error (two-sided) of 5%. However, such information is currently not available for primary progressive multiple sclerosis.

VI.2.G. Study population
Patients with primary progressive multiple sclerosis according to accepted criteria and age 18-55 years should be studied. It is a problem that patients with primary progressive multiple sclerosis do not readily conform to accepted criteria (McDonald criteria or Poser criteria) and have a wide differential diagnosis. Only patients with moderate disability (EDSS 3 to 5.5) should be included. Enrolled patients should have suffered recent clinical disease activity, i.e. progression or relapses during the last 1-2 years.

VI.2.H. Specific inclusion criteria
Similar to those described for phase II placebo-controlled trial.
VI.2.I. Specific exclusion criteria
Similar to those described for phase II placebo-controlled trial.

VI.2.J. Data analysis methods
The analysis of the primary efficacy variable is based on the intention-to-treat (ITT) population. Time to progression is analysed using the log-rank test and Kaplan-Meier estimates.

VII. OTHER STUDIES (ATYPICAL MS FORMS)

Rare inflammatory demyelinating syndromes, such as Marburg’ Disease, Pseudotumoral forms, Baló’s Concentric Sclerosis (BCS) or Devic’s Neuromyelitis Optica (DNO), are difficult to be studied in a randomised controlled clinical trial mainly due both to absence of specific clinical and laboratory findings leading to an early and definitive diagnosis and to the small number of patients seen every year with homogeneous clinical characteristics (for example at the onset of disease, after a few clinical relapses and low disability). As a consequence, to date no reliable data are available on the efficacy of immunomodulatory or immunosuppressive drugs, usually used in MS, in halting or slowing the inflammatory and degenerative processes underlying these rare inflammatory diseases.

VII.1. Marburg’s disease
Marburg’s disease is an acute malignant monophasic demyelinating disease, which usually results in death within a few weeks or months after the initial bout. It is characterized by widespread and progressive cerebral white matter destruction or by severe pathological involvement of clinically strategic regions such as brainstem, resulting in bulbar paralysis. Short-term, observational studies of a restricted cohort of patients evaluating the therapeutic efficacy of the association of high dosage Metilprednisolone and frequent, pulse administration of immunosuppressive drugs (cyclophosphamide or mitoxantrone) or plasmapheresis (with or without subsequent pulse cyclophosphamide) have been reported.

VII.2. Baló’s Concentric Sclerosis and Pseudotumoral forms
Baló’s Concentric Sclerosis is a rare demyelinating disorder characterized pathologically by concentric rings of alternating demyelinated and relatively myelin preserved white matter. The pathogenesis of the concentric lesion may be explained by periodic suppression of demyelination in a rapidly expanding area of inflammation, allowing remyelination or only transient incomplete demyelination to occur. While initial reports of BCS predicted that the disease was rapidly progressive and fatal, a good prognosis has been recently described in a few cases and the lesion is usually considered as an atypical, large, pseudotumoral MS plaque. Large, focal, tumor-like demyelinating lesions of the CNS often represent a diagnostic challenge, which reasonably calls for a stereotactic biopsy, particularly when isolate in the brain, to exclude glioma, infectious processes or primary CNS vasculitis. These cases are usually characterized by a severe course and a rapid clinical deterioration. Response to therapy is highly variable. Patients are usually treated with steroid, generally in high doses intravenously, as well as variably with cyclophosphamide or plasma exchange. Despite the aggressive therapy sometimes the disease ultimately progresses, mainly in patients with spinal cord involvement.

VII.3. Devic’s Neuromyelitis Optica (DNO)
DNO is usually considered as a distinct disease entity from Ms according to several clinical, neuroradiological and CSF findings. Pathological aspects and the relevant pathogenetic role of humoral immunity also support the belief that DNO may be a separate syndrome. DNO diagnosis is difficult to make after the first episode. Two or more acute episodes of neurological impairment involving the optic nerves and the spinal cord, in a simultaneous or sequential temporal relationship, must be observed, with no clinical or MRI evidence of brain involvement. A poor prognosis may be predicted by the age at onset, the interval between first and second episode, the relapse rate and the severity of the first attack. At the moment no effective treatment has been demonstrated, although chronic or pulse steroid therapy is often used in DNO
patients, associated with a variety of immunosuppressive or immunomodulatory agents, empirically chosen, usually after 3-5 attacks have occurred.

VII.4. Future therapeutic strategies
New emerging disease-modifying therapies that target cytokines, the blood brain barrier, the “trimolecular complex” or that act by deletion of auto-reactive T cells are candidates for treatment of these rare MS variants and their efficacy and safety should also be evaluated. These syndromes are characterised by a very aggressive course, therefore the research of active and efficacious treatments needs to be strongly encouraged. Since most of these patients have usually a bad prognosis and deteriorate to severe irreversible disability in only a few months or years, beneficial effect due to therapeutic intervention may be easily detected by using strong primary end points, such as death or loss of deambulation. Classical phase II-III studies are impossible in this MS variants; small group of patients treated and monitored with a specific protocol are the only feasible approach, even if it implies problems with interpretation and generalization of the results.

The Restricted Cohort Study applies the principles and patient enrolment procedures regularly used in randomised clinical trials. Therefore, strict eligibility criteria and the appropriate choice of zero time must be well defined; anyway baseline differences should be adjusted for prognostic risk. In addition, patients must be classified according to suitable clinical criteria to enable adjustment for any inequalities in susceptibility to the outcome. Finally, the analysis of the data should be conducted using the same methodology used in clinical trials.

The primary objective of these studies is to provide preliminary data on the efficacy, considered both on short term evolution of the recent clinical attack as well as on long term evolution of the disease, evaluated as confirmed changes of EDSS and ambulation index. MRI and neurophysiological test are very important to support clinical observation. The secondary objectives of these studies are to gather descriptive information concerning short and long term tolerability and safety of the investigational therapeutic intervention as well as about the potential loss of efficacy on long-term chronic or pulse administration.

VIII. EXAMPLES OF PHASE III TRIALS IN MULTIPLE SCLEROSIS

Pivotal placebo-controlled trials in relapsing remitting multiple sclerosis
6. Comi G, Filippi M, Wolinsky JS. European/Canadian multicenter, double-blind, randomized, placebo-controlled study of the effects of glatiramer acetate on magnetic resonance imaging--measured disease

**Comparative superiority Phase III trials in relapsing remitting multiple sclerosis**

**Phase III placebo-controlled trial in early multiple sclerosis**

**Phase III placebo-controlled trials in secondary multiple sclerosis**

**IX. SUGGESTED READINGS**
Pharmacological Research in Mental Disorders

Chapter 21. Mood Disorders

Chapter 22. Anxiety Disorders

Chapter 23. Schizophrenic Disorders

Chapter 24. Alcoholism and Nicotine Addiction

GLOSSARY OF TERMS

<table>
<thead>
<tr>
<th>Acronym</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>ADS</td>
<td>Alcohol Dependence Scale</td>
</tr>
<tr>
<td>AIMS</td>
<td>Abnormal Voluntary Movement Scale</td>
</tr>
<tr>
<td>ARCI</td>
<td>Addiction Research Center Inventory</td>
</tr>
<tr>
<td>BAS</td>
<td>Barnes Akathisia Scale</td>
</tr>
<tr>
<td>BDI</td>
<td>Bipolar I Disorder</td>
</tr>
<tr>
<td>BDI</td>
<td>Beck Depression Inventory</td>
</tr>
<tr>
<td>BSRS</td>
<td>Brief Psychiatric Scale</td>
</tr>
<tr>
<td>CGI</td>
<td>Clinical Global Impression Scale</td>
</tr>
<tr>
<td>CIWA</td>
<td>Clinical Institute Withdrawal Assessment for Alcohol</td>
</tr>
<tr>
<td>DRUG AAA</td>
<td>Investigational Drug for Treatment of Alcohol Dependence</td>
</tr>
<tr>
<td>DRUG SSS</td>
<td>Investigational Drug for Treatment of Nicotine Dependence</td>
</tr>
<tr>
<td>DRUG XOXO</td>
<td>Investigational Drug for Treatment of MDD</td>
</tr>
<tr>
<td>DRUG XXX</td>
<td>Investigational Drug for Treatment of Anxiety</td>
</tr>
<tr>
<td>DRUG YYY</td>
<td>Investigational Drug for Treatment of Schizophrenia</td>
</tr>
<tr>
<td>DSM-IV</td>
<td>Standard Diagnostic Criteria for Mental Disorders</td>
</tr>
<tr>
<td>ESRS</td>
<td>Extrapyramidal Symptom Rating Scale</td>
</tr>
<tr>
<td>FTND</td>
<td>Fagerstrom Test for Nicotine Dependence</td>
</tr>
<tr>
<td>GAD</td>
<td>General Anxiety Disorder</td>
</tr>
<tr>
<td>HAM-A</td>
<td>Hamilton Rating Scale for Anxiety</td>
</tr>
<tr>
<td>HAM-D</td>
<td>Hamilton Rating Scale for Depression</td>
</tr>
<tr>
<td>MADRS</td>
<td>Montgomery Asberg Depression Scale</td>
</tr>
<tr>
<td>MAOI</td>
<td>Monoamine Oxidase Inhibitors</td>
</tr>
<tr>
<td>MAST</td>
<td>Michigan Alcohol Screening Test</td>
</tr>
<tr>
<td>MDD</td>
<td>Major Depressive Disorder</td>
</tr>
<tr>
<td>OBS</td>
<td>Obsessive Compulsive Syndrome</td>
</tr>
<tr>
<td>PANSS</td>
<td>Positive and Negative Symptom Scale</td>
</tr>
<tr>
<td>PD</td>
<td>Panic Disorder</td>
</tr>
<tr>
<td>POMS</td>
<td>Profile of Mood States</td>
</tr>
<tr>
<td>SAS</td>
<td>Simpson Angus Scale</td>
</tr>
<tr>
<td>SHAPS</td>
<td>Snaith-Hamilton Pleasure Scale</td>
</tr>
<tr>
<td>SNRI</td>
<td>Serotonin-Norepinephrine Re-uptake Inhibitors</td>
</tr>
<tr>
<td>SSRI</td>
<td>Selective Serotonin Re-uptake Inhibitors</td>
</tr>
<tr>
<td>STA-IX</td>
<td>Stait-Trait Anxiety Scale</td>
</tr>
<tr>
<td>TCA</td>
<td>Tricyclic Antidepressants</td>
</tr>
<tr>
<td>VAAS</td>
<td>Visual Analogue Anxiety Scale</td>
</tr>
<tr>
<td>VAS</td>
<td>Visual Analogue Scale</td>
</tr>
</tbody>
</table>
Chapter 21. Mood Disorders

Claudio A. Naranjo, M.D. ¹

Lara Chayab ²

¹ Professor
Department of Pharmacology
Psychiatry and Medicine
University of Toronto
Toronto, Ontario
CANADA
and
Head
Neuropsychopharmacology Research Program
Sunnybrook & Women’s College Health Science Centre
Toronto, Ontario
CANADA

² M.Sc. Candidate
Department of Pharmacology
University of Toronto
Toronto, Ontario
CANADA
and
Researcher
Neuropsychopharmacology Research Program
Sunnybrook & Women’s College Health Science Centre
Toronto, Ontario
CANADA
I. INTRODUCTORY REMARKS

The mood disorders that have been most extensively examined in epidemiological studies are major depressive disorder (MDD), dysthymia, and bipolar I disorder (BDI) (1). MDD is the most prevalent of the other disorders. The lifetime prevalence rates of MDD show that 5%-12% of men will experience depression at some point in their lifetime while that rate is higher in women accounting for 10%-25% (2). The diagnostic criteria of MDD according to the DSM-IV (American Psychiatric Association, 1994) are the feeling of sadness and/or loss of pleasure present most of the day, everyday, for at least two weeks (anhedonia). During this period, at least five other symptoms must be present including appetite disturbances, weight disturbances, sleep disturbances, activity disturbances, fatigue, inappropriate guilt, and thoughts of death (3). The lifetime prevalence rate of Bipolar I Disorder is much less than the prevalence of MDD accounting for only 1 to 2% (2). However, individuals exhibiting BDI experience more depressive episodes than those with MDD do (2). The diagnostic criteria for BDI according to the DSM-IV are the occurrence of one or more manic episodes or mixed episodes. Often individuals have also had one or more Major Depressive Episodes. Pharmacotherapy for BDI includes mood stabilizers, such as lithium and valproate, as first line of treatment and mood stabilizers with other medications for people unresponsive to first line treatments (1). During a manic episode, individuals experience hyperactivity, hallucinations, and paranoia (2). These symptoms usually cause problems to the diagnosed patients with the law, at work, and with other individuals. Because of the nature of the disorder, clinical studies aimed at investigating treatments for BDI are difficult to establish because of the lack of compliance and the large dropout rate.

Choices of pharmacological therapy for the treatment of depression include tricyclic antidepressants (TCAs), monoamine oxidase inhibitors (MAOIs), and newer agents such as selective serotonin reuptake inhibitors (SSRIs), and serotonin-norepinephrine reuptake inhibitors (SNRI) (1). TCAs and MAOIs have been associated with severe adverse effects, drug interactions, and toxicities. Side effects of TCAs administration include dry mouth, constipation, blurred vision, sedation, weight gain, and sexual dysfunction while the most frequent adverse effects of MAOIs are similar to TCAs and also include orthostatic hypertension, palpitations, tachycardia, peripheral edema, and muscle cramps (2). Depressed patients usually terminate the drug therapy due to the side effects before the full course of the treatment is achieved, which leads to recurrence of symptoms. SSRIs and SNRIs have a better profile of adverse effects. They are less sedating, and have no cardiac effects. However, they still cause nausea, headaches, insomnia, sexual dysfunction, tremors, and CYP 450 inhibition (2). If lower doses of the drugs are used to minimize adverse effects, the efficacy of the drugs will be negatively affected and full recovery will not be achieved. In addition, approximately 5 weeks are required for the onset of action of SSRIs, and sometimes about 1 week for the onset of action of SNRIs (2). As a result, there is a need for newer antidepressant agents that have great efficacy for moderate to severe depression, better profile of adverse effects, fewer drug interactions, and a faster onset of action.

In this chapter, a double-blind, placebo-controlled, parallel-group study design will be used to test the efficacy and safety of investigational drug XOXO for patients with moderate to severe depression. Previous clinical studies have shown that drug XOXO is a highly selective serotonin reuptake inhibitor, is linear and dose-proportional, has a half-life of 27-32 hrs which accounts for its once daily dosing, its onset of action occurs within one week, and is eliminated by biotransformation. In addition, drug XOXO has a low potential for drug-drug interactions since its effect on CYP 450 is negligible, and has a favorable profile of adverse effects. In this study design, these properties of the drug will be better characterized and its efficacy will be determined. The example of the study design illustrated in this chapter is most applicable to trials for evaluating antidepressants intended to be used in patients as first line of treatment.
II. PHASE II STUDIES FOR REGISTRATION OF NEW ANTIDEPRESSANT DRUGS

II.1. Outline of a typical development plan

This study will examine the efficacy and safety of drug XOXO in men and women experiencing moderate to severe depression. All patients enrolled in the study meeting inclusion/exclusion criteria and that give consent will be randomly assigned to receive one oral daily dose of drug XOXO or placebo for 12 weeks. Efficacy and safety measures are going to be performed at weekly intervals up to the end of week 12. The study will be a randomized, double blind, placebo-controlled and patients will come back for two follow-up assessments.

II.2. Short-term studies

II.2.a. Study Objectives

Primary Objectives
a. To compare the efficacy of drug XOXO treatment versus placebo in reducing the symptoms of MDD
b. To compare the safety of drug XOXO treatment versus placebo

Secondary Objectives
a. To determine the onset of the antidepressant action of drug XOXO
b. To determine the duration of the antidepressant effect of drug XOXO
c. To determine the peak antidepressant effects of drug XOXO

II.2.b. Primary Endpoints

The efficacy of antidepressant drugs in clinical trials is measured using a wide variety of assessment tools, which include clinical observations, interviews, and self-reports. Currently, a number of rating scales exists that provide a standardized approach to evaluate the severity of mental disorders and the treatment outcomes. Scales are designed to measure either general symptoms or disease-specific symptoms. Some scales have to be rated by psychiatrists; nurses or research assistants can rate others, and yet other scales are self-evaluated. The choice of the appropriate scales for the diagnosis of specific mental disorder and the evaluation of the efficacy of investigational drugs depends on the specificity, sensitivity, and simplicity of the scales in question. The assessment tools have been extensively researched and evaluated for their specificity and sensitivity for each of the mental disorders that are going to be discussed in this chapter.

Rating scales will be administered to assess the following dependent variables:

a. Structured Clinical Interview for DSM-IV to diagnose patients with major depressive disorder
b. The Montgomery Asberg Depression Scale, MADRS, (Score ≥ 30 for severely depressed patients) and The Beck Depression Inventory, BDI, (Score ≥ 16 for severely depressed patients) to assess current level of depression (4).
c. The Hamilton Depression Scale, HAM-D to assess severity of depressed mood (score ≥ 17 for severely depressed patients), which contains 17 items to assess depressed mood, suicidal ideation, somatic symptoms, and loss of interest. Four additional items are included (i.e. diurnal variation, derealization, paranoid symptoms and obsessional symptoms), making the total questionnaire 21 questions in length (4).
d. The Beck Anxiety Inventory and/ Stait-Trait Anxiety, STA-IX, to exclude patients with comorbid depression and anxiety, which is composed of 21 questions and evaluates the current level of depression (4).
Responders to the antidepressant drug XOXO versus placebo, where a response is defined as:

a. \( R = \) a reduction from baseline (weeks 1) on weeks 2, 4, 6, 8, 10, 12 during and post treatment as measured by the HAM-D.

b. \( R = \) a reduction from baseline (week 1) on weeks 2, 4, 6, 8, 10, 12 during and post treatment as measured by the BDI.

c. \( R < 2 \) on weeks 2,4,6,8,10,12 during and post treatment, as measured by the Snaith-Hamilton Pleasure Scale (SHAPS), which is a validated self-assessment scale estimating the degree to which a person is able to experience pleasure or the anticipation of a pleasurable event (i.e. hedonic tone). A score of 2 or more "disagree/definitely disagree" is considered to be indicative of an anhedonic state (4).

d. \( R = \) a reduction in the score, on weeks 2,4,6,8,10,12 during and post treatment, as measured by the Addiction Research Center Inventory (ARCI), which is a 77-item questionnaire that measures subjective effects of drugs’ positive/reinforcing (e.g. euphoria, stimulation) and negative or dysphoric (e.g. sedation, confusion). This inventory allows the quantification of subjective drug effects with scales sensitive to the effects of specific drugs and drug classes (4).

e. \( R = \) a reduction in the score, on weeks 2,4,6,8,10,12 during and post treatment, as measured by the Profile of Mood States (POMS), which is commonly used for assessing drug-induced changes in mood, the POMS consists of a series of 72 adjectives. With respect to each adjective, subjects respond how they feel using a five-point scale ranging from "extremely" to "not at all". Tension-Anxiety, Anger-Hostility, Depression-Dejection, Friendliness, Fatigue, Confusion, Vigor, Elation, Arousal, and Positive Mood are the 10 scales covered in the POMS (4).

II.2.c. Secondary Endpoints

a. The time of onset of a consistent decrease in depressed mood as measured by the Visual Analogue Scale (VAS) compared to baseline. The VAS is often used in the assessment of momentary changes in affect. They consist of a selection of visual analog rating scales (100mm lines) anchored at each end by opposing adjectives to evaluate drug "liking", drug effect and desire to experience the drug effects again. Subjects are instructed to rate how they feel by making a mark anywhere along the line (4).

b. The treatment day during which the greatest reduction in mood is present as measured by the HAM-D.

c. The number of patients that achieve HAM-D \( \leq 7 \).

d. Adverse effects are measured by changes in blood pressure, heart rate, GI motility, and abnormal laboratory tests (blood and liver).

II.2.d. Study Design

This is a double blind, randomized, placebo-controlled, parallel-group study. There will be 2 groups in this study, a moderate to severely depressed group, and a healthy control group. Patients in each group will receive either a single dose of drug XOXO or a single dose of placebo randomly once daily for 12 weeks. The number of patients receiving drug XOXO will equal the number of patients receiving placebo within each group. The antidepressant effects of drug XOXO will be measured using various questionnaires (discussed below). Objective measures, such as blood pressure, will also be obtained. The results will be compared and analyzed between and within groups.

Screening for Eligibility (Visit 1, Week 1)

Subjects will be interviewed to ensure suitability for study participation a week before the commencement of the study. The following will be required to assess eligibility:

a. Written informed consent.

b. Structured Clinical Interview for DSM-IV to assess psychiatric status and to rule out dependence on psychoactive substances.
c. Current level of depression (Hamilton Depression Scale (HAM-D), Montgomery Asberg Depression Scale (MADRS), and Beck Depression Inventory (BDI)), and anxiety (Beck Anxiety Inventory and/ State-Trait Anxiety (STA-IX)). This is the baseline measure to which all upcoming results will be compared against.

d. Brief medical examination (heart rate, blood pressure).

e. Medical history.

f. Review of inclusion/exclusion criteria.

g. Pregnancy test for women.

h. Blood and urine samples to assess liver function, hematology, biochemistry, and to detect the presence of other psychoactive drugs.

i. Participants will be provided with a pager number to be used if they experience any serious side effects and a wallet card containing information about participation in this study.

j. Patients are also given a diary in which they record drug compliance and any side effects that they may experience on a daily basis.

Treatment Phase (Visits 2 – 7, Weeks 2-12)

a. Eligible subjects will attend six treatment sessions (one every two weeks).

b. Medical examination.

c. Medication will be dispensed (enough pills for two weeks).

d. Treatment will take place at 2-week intervals consisting of 30 to 45 minute sessions with the research assistant.

e. A psychiatrist will be available for consultation, assessment, and treatment as needed (i.e. adverse drug reaction, increases in severity of depressive symptomatology).

f. Review Daily Diary forms on which patients record compliance with medication.

g. At each visit, the MADRS, BDI, HAM-D, SHAPS, ARCI, POMS, and VAS will be completed and subjects will be interviewed regarding concomitant illness and medication use.

h. Ask patients to return any unused medication in the vial.

i. Blood will be drawn for trough drug concentrations at visits 3, 5 and 7 (4, 8 and 12 weeks after commencing medication).

j. Blood and urine will also be collected at visits 4, 6 and 8 for drug screen, complete biochemistry and hematology analysis.

k. Subjects will be referred to their family physicians either at the end of the 12 week study or if a subject decides to terminate participation in the study.

l. Individuals who do not respond to drug XOXO will be referred to alternate psychiatric treatment or to their family physicians.

Follow-up visits (Visits 8-9, at 3 months and 6 months after treatment)

a. Review daily diary

b. Medical examination

c. Psychiatrist: examine any increased depressed symptoms, interview patients for concomitant illness and examine potential adverse reactions.

d. Complete questionnaires: MADRS, BDI, STAI-XI, HAM-D, SHAPS, ARCI, POMS, and VAS.

e. Blood and urine collection for drug screen, complete biochemistry, and hematology analyses.

II.2.e. Planned Sample

Flemming’s Single Stage Procedure will be used in calculating the sample size in the demonstrated phase II study (5). The procedure depends on the assumption that investigators usually have some knowledge of the activity of drugs similar to the one being studied. Therefore, in this study, researchers will compare the anticipated response of drug XOXO to other observed responses of similar drugs with the same
therapeutic indication. Researchers will then specify a probability of a response, which could then be compared to the actual responses to standard treatments. If the response exceeds that of standard treatments, then it can be concluded that Drug XOXO exhibits efficacy (5).

Therefore, assume:
Largest response proportion = Ro
Smallest response proportion = Ra
Hypothesis:

\[
R \leq Ro \\
R \geq Ra
\]

\(\alpha\), probability of rejecting the hypothesis of \(R \leq Ro\)
\(\beta\), probability of rejecting the hypothesis of \(R \geq Ra\)

For \(N\) patients recruited for phase II trial, the observed number of patient responses \(r\) has a binomial distribution with parameter \(\pi\).

Therefore, the sample size required for Flemming’s Single Stage Procedure is approximately,

\[
N = \left[ Z_{1-\alpha} \sqrt{Ro(1-Ro)} + Z_{1-\beta} \sqrt{Ra(1-Ra)} \right]^2 / [Ra-Ro]^2
\]

A treatment regimen using SSRIs indicated in the treatment of MDD in phase II studies is expected to yield a response in at least 35% of the patients being tested to show efficacy. Previous phase I trials have shown that Drug XOXO exhibits a higher safety profile than standard treatments indicated for MDD. A one-sided test size is set at 5% and the power at 80%. Since the new investigational drug is shown to be safer than the standard treatments, the values of \(Ro\) and \(Ra\) will be set at 0.6 and 0.5 respectively with \(\alpha = 0.05\) and \(1-\beta = 0.8\). Therefore, from Table 12.1 in the “Statistical Tables for the Design of Clinical trials” handbook, or from calculating the equation, the sample size \(N\) will equal 16 (5). Therefore, at least 16 patients are needed in each group (16 in the MDD group and 16 in the healthy control group) to detect significance in efficacy for this trial. Therefore, a total of 38 patients are going to be enrolled for the successful completion of this study.

II.2.f. Study Population
Male or Female over 18 years of age meeting DSM-IV criteria for MDD and who exhibit moderate to severe depression or a Ham-D score of \(\geq 17\).

II.2.g. Specific Inclusion Criteria
A subject will be eligible for inclusion in the study only if all of the following criteria apply:

- Males or females between 19 to 50 years of age.
- Socially stable.
- Meet DSM-IV criteria for major depressive disorder.
- In-patients or out-patients.
- Non-smokers.

II.2.h. Specific Exclusion Criteria
Exclusion criteria must take into account the characteristics of the drug (pharmacokinetics, pharmacodynamics, drug-drug interactions, and adverse effects). Patients will not be eligible to participate in the study if one of the following criteria apply:

- If meet criteria for Bipolar Disorder, schizophrenia, schizo-affective or other substance abuse/dependence.
- Evidence of medical or surgical illness requiring treatment.
- History of psychoactive drug dependence or a positive urine test for psychoactive drugs.
- Use of medications which may interfere with the study procedures (e.g. SSRIs).
e. Any clinically significant abnormality evident in biochemistry or hematology test results or in urine analysis requiring further investigation.

f. Active suicidal ideation.

g. Receiving or will receive other investigational drug during the study.

h. Pregnant or lactating females.

II.2.i. Tools to assess endpoints

Efficacy should be evaluated using the tools depicted in the following table:

<table>
<thead>
<tr>
<th>Variable</th>
<th>Assessment Tools to Measure Variable</th>
<th>Time of Assessment</th>
</tr>
</thead>
<tbody>
<tr>
<td>MDD diagnosis</td>
<td>DSM-IV, HAM-D</td>
<td>Visit 1, Week 1</td>
</tr>
<tr>
<td>Level and severity of depression</td>
<td>HAM-D, MADRS, BDI</td>
<td>Visit 1, Week 1</td>
</tr>
<tr>
<td>Excluding patients with concomitant anxiety</td>
<td>STA-XI</td>
<td>Visit 1, Week 1</td>
</tr>
<tr>
<td>Reduction in depressed mood</td>
<td>HAM-D, POMS</td>
<td>Visits 2 - 9</td>
</tr>
<tr>
<td>Reduction in loss of interest (ability to experience pleasure)</td>
<td>HAM-D, SHAPS</td>
<td>Visits 2 - 9</td>
</tr>
<tr>
<td>Increase in Euphoria</td>
<td>HAM-D, ARCI</td>
<td>Visits 2 - 9</td>
</tr>
<tr>
<td>Decrease in dysphoria</td>
<td>ARCI</td>
<td>Visits 2 - 9</td>
</tr>
<tr>
<td>Increase/decrease in hostility, fatigue, and confusion</td>
<td>POMS</td>
<td>Visits 2 - 9</td>
</tr>
<tr>
<td>Time of onset of consistent decrease in depressed mood</td>
<td>VAS, HAM-D</td>
<td>Visits 2 - 9</td>
</tr>
<tr>
<td>Time of greatest reduction in depressed mood</td>
<td>HAM-D</td>
<td>Visits 2 - 9</td>
</tr>
<tr>
<td>Number of patients achieving HAM-D ≤ 7 after end of study</td>
<td>HAM-D</td>
<td>Visits 2 - 9</td>
</tr>
<tr>
<td>Drug Compliance</td>
<td>Daily Diary and returned pills</td>
<td>Visits 2 - 7</td>
</tr>
</tbody>
</table>

Tools to assess safety

Adverse events such as GI abnormalities, blood pressure and heart rate changes, and blood biochemistry and hematology changes will be assessed in this study.

a. A complete medical examination will be performed, by a physician, at baseline (visit 1) as well as during all visit days. Any changes in blood pressure, heart rate, GI motility, and other complaints made by the patient will be recorded and compared to baseline. In the case of a patient developing any kind of adverse reaction, the subject will be immediately asked to return all medications and withdraw from the study. The patient will then be referred to his/her family doctor to avoid further complications.

b. Urine and blood tests will be performed on visits 2, 4, 6, 8, and 9. Any changes will also be reported and the patient will be asked to withdraw from the study.

c. The daily diaries are provided for the patients to record their feelings, drug compliance, and the occurrence of any adverse effects daily. The diaries will then be reviewed by the psychiatrist and their contents discussed by the patients.

d. If a serious side effect develops in a patient, a full analysis will be made to ensure that the adverse effect is from the investigational drug and not caused by other drugs that the patient may have taken, drug interaction, or a disease.

e. The adverse effects caused by the investigational drug XOXO and those developing from placebo will be compared to determine if a significant difference exists in order to identify the safety profile of drug XOXO.
II.2.j. Specific criteria for early withdrawal and discontinuation

Subjects are allowed to withdraw from the study at any time. Subjects must leave the study if one of these conditions holds:

a. Occurrence of serious side effects
b. Pregnancy
c. Non-compliance
d. Development of a medical condition
e. Use of other medication
f. Violation of the protocol
g. Withdrawal of consent

Subjects that terminate their involvement in the study because of the occurrence of side effects will be considered as having completed the study. Blood tests, urine tests and a complete medical exam should be performed on these patients. The results should determine whether the patients need to be placed on therapy to eliminate the side effects or whether the side effects will resolve on their own. In addition, these patients will be referred to their family physician and an assessment session should be performed after 3 months of withdrawal. Subjects that withdraw because of use of other medications, non-compliance, violation of the protocol, pregnancy, or development of a medical condition will be considered as having not completed the study and will be replaced by new participants. Subjects that terminate the study after drug administration should be contacted and followed-up to ensure that no severe side effects or worsening of the condition takes place. These subjects should also be referred to their family physician and called for a follow-up assessment after 3 months.

II.2.k. Data Analysis Method

For each visit, subject group (ex. Depressed vs. Controls) and dependent variables (e.g. HAM-D scores), parameters such as the mean, maximum, minimum change from baseline will be calculated and analyzed for the effects of drug XOXO. An analysis of variance will be conducted to compare all groups in order to determine if there is a significant difference in the way the groups responded to drug XOXO challenge. These data will also be entered into an ANOVA in order to evaluate the role of depression on the effect of drug XOXO. Together, these analyses should provide information on the efficacy and safety of drug XOXO on patients with MDD.

III. REFERENCES

Chapter 22. Anxiety Disorders

Claudio A. Naranjo, M.D. ¹

Lara Chayab ²

¹ Professor
Department of Pharmacology
Psychiatry and Medicine
University of Toronto
Toronto, Ontario
CANADA
and
Head
Neuropsychopharmacology Research Program
Sunnybrook & Women’s College Health Science Centre
Toronto, Ontario
CANADA

² M.Sc. Candidate
Department of Pharmacology
University of Toronto
Toronto, Ontario
CANADA
and
Researcher
Neuropsychopharmacology Research Program
Sunnybrook & Women’s College Health Science Centre
Toronto, Ontario
CANADA
I. INTRODUCTORY REMARKS

Anxiety Disorders are the most common forms of mental illness (1). The National Comorbidity Survey indicated that about 48% of the US population of Americans aged 15 to 54 years had at least one mental illness in their lifetime. Of these, 24.9% reported having an anxiety disorder in their lifetime (1). Anxiety disorders include general anxiety disorder (GAD), panic disorder (PD), and obsessive compulsive disorder (OBC) (2). Of the three disorders, GAD has been shown to be the most prevalent in a population. Patients are required to exhibit at least six symptoms of hyperarousal, vigilance, motor tension, and autonomic hypersensitivity to fit the criteria of GAD in the DSM-IV (2). Pharmacotherapies of anxiety disorders include treatment with benzodiazepines, buspirone, and antidepressants (3). Benzodiazepines have been known for their high efficacy and their fast onset of action. However, they are only recommended for treatment of acute anxiety as long term use of benzodiazepines leads to a wide range of adverse effects including sedation, improper coordination, memory loss, depression, dependence, and potential for abuse (3). The use of antidepressants, venlafaxine, imipramine, and paroxetine, in the treatment of anxiety has been shown to be efficacious; however, these drugs exhibit a delayed onset of action and various adverse effects including nausea, insomnia, jitteriness, restlessness and agitation. As a result, patients on antidepressants usually terminate the treatment before full recovery from anxiety is accomplished (3). The use of TCAs, such as clomipramine, is also efficacious; however, its anticholinergic side effects are so severe that patients also tend to end the treatment before full recovery is reached (1). Studies with buspirone have showed that it also exhibits a more gradual onset of action and may have a lower efficacy than benzodiazepines. Its adverse effects include GI symptoms and dizziness (3).

As a result, a newer anxiolytic agent is required that has similar or better characteristics of efficacy and onset of action as benzodiazepines, exhibits a better profile of safety and no dependence, and indicated for long term treatment of anxiety.

In this chapter, a double blind, placebo controlled, parallel-group study design will be used to test the efficacy and safety of investigational drug XXX versus placebo in patients diagnosed with moderate to severe anxiety. Previous clinical studies have shown that drug XXX is a highly potent serotonin 1A receptor agonist, is not structurally or chemically related to benzodiazepines, has no sedative effect, does not lead to dependence, has a fast onset of action, requires once daily dosing and is eliminated by biotransformation. In this study design, these properties of the drug will be better characterized and its efficacy will be determined. The example of the study design illustrated in this chapter is most applicable to trials for evaluating anxiolytic drugs intended to be used in patients with moderate to severe anxiety as the first line of treatment.

II. PHASE II STUDIES FOR REGISTRATION OF NEW ANXIOLYTIC DRUGS

II.1. Outline of a typical development plan

This study will examine the efficacy and safety of drug XXX in men and women experiencing moderate to severe anxiety. All patients enrolled in the study meeting inclusion/exclusion criteria and that give consent will be randomly assigned to receive one oral daily dose of drug XXX or placebo for 6 months. Efficacy and safety measures are going to be performed at weekly intervals up to week 4, then at 2-week intervals up to week 12, and then at 4-week intervals up to the end of the 6 months. The study will be randomized, double blind, placebo-controlled and patients will come back for two follow-up assessments.

II.2. Long-term studies

II.2.a. Study Objectives
Primary Objectives
   a. To compare the efficacy of drug XXX treatment versus placebo in reducing the symptoms of anxiety.
   b. To compare the safety of drug XXX treatment versus placebo
Secondary Objectives
a. To determine the onset of the anxiolytic action of drug XXX
b. To determine the duration of the anxiolytic effect of drug XXX
c. To determine the peak anxiolytic effects of drug XXX

II.2.b. Primary Endpoints
Rating scales will be administered to assess the following dependent variables:

a. Structured Clinical Interview for DSM-IV to diagnose patients with Generalized Anxiety Disorder
b. The Hamilton Rating Scale for Anxiety, HAM-A to assess severity of anxiety (score $\geq 18$ for patients with severe anxiety and a score of $\geq 2$ on the HAM-A item 1 (anxious mood) and item 2 (tension)). It is a simple 14-item five step rating scale. Each item is a group of symptoms that represents one general criterion associated with the disorder. For example, the combination of worries, anticipation of the worst, apprehension, and irritability determine anxious mood. This scale is simple, specific for anxiety measurement, and sensitive to drug effects (4).
c. The Stait-Trait Anxiety, STA-IX to measure subject selection (the Trait) and treatment effects (the Stait). It is a self-evaluation scale that consists of 20 items with a four step severity scale. The Questionnaire asks the patient to indicate how he/she feels at the moment (state) or in general (trait) by selecting “not at all”, “somewhat”, “moderately so”, or “very much so”, with the latter indicating high anxiety (4).

Responders to the anxiolytic drug XXX versus placebo, where a response is defined as:

$R =$ a reduction from baseline (week 1) to week 8 and a similar or a further reduction from baseline to the end of month 6 as measured by the HAM-A, in order to determine the differences in efficacy and safety between the short-term outcome (by week 8) and the long-term outcome (by the end of month 6).

$R =$ a reduction in anxiety measures from baseline to week 8 and a similar or further reduction from baseline to the end of month 6 as measured by the STA-IX scale.

II.2.c. Secondary Endpoints

a. The time of onset of a consistent decrease in anxious mood as measured by the Visual Analogue Anxiety Scale (VAAS) compared to baseline (4).
b. The day during which the greatest reduction in anxious mood is present as measured by the HAM-A.
c. The number of patients that achieve HAM-A $\leq 10$ or full recovery.

Adverse effects are measured by changes in blood pressure, heart rate, GI motility, and abnormal laboratory tests (blood and liver).

II.2.d. Study Design
This is a double blind, randomized, placebo-controlled, parallel-group study. The subjects will be administered either a single oral dose of drug XXX or a single oral dose of placebo daily for 6 months. The number of patients receiving drug XXX will equal the number of patients receiving placebo within each group. The anxiolytic effects of drug XXX will be measured using various questionnaires (discussed below). Objective measures, such as blood pressure, will also be obtained. The results will be compared and analyzed between and within groups.

Screening for Eligibility (Visit 1)
Subjects will be interviewed to ensure suitability for study participation a week before the commencement of the study. The following will be required to assess eligibility:

a. Written informed consent.
b. Structured Clinical Interview for DSM-IV to assess psychiatric status and to rule out dependence on psychoactive substances.
c. Current level of anxiety (Hamilton Depression Scale (HAM-A), State-Trait Anxiety (STA-IX)). This is the baseline measure which all upcoming results will be compared against.
d. Brief medical examination (heart rate, blood pressure).
e. Medical history.
f. Review of inclusion/exclusion criteria.
g. Pregnancy test for women.
h. Blood and urine samples to assess liver function, hematology, biochemistry, and to detect the presence of other psychoactive drugs.
i. Participants will be provided with a pager number to be used if they experience any serious side effects and a wallet card containing information about participation in this study.
j. Patients are also given a diary in which they record drug compliance and any side effects that they may experience on a daily basis.

**Treatment Phase (Visits 1-10, Months 1-6)**

a. Eligible subjects will attend 11 treatment sessions (weekly intervals up to week 4, then at 2-week intervals up to week 12, and then at 4-week intervals up to the end of the 6 months).
b. Medical examination.
c. Medication will be dispensed (enough pills to last for the next visit at once daily dosing).
d. Treatment will consist of 30 to 45 minute sessions with the research assistant.
e. A psychiatrist will be available for consultation, assessment, and treatment as needed (i.e. adverse drug reaction, increases in severity of anxious symptomatology).
f. Review Daily Diary forms on which patients record compliance with medication and any side effects that the patient may be experiencing.
g. At each visit, the HAM-A and the STA-IX, and VAS will be completed and subjects will be interviewed regarding concomitant illness and medication use.
h. Ask patients to return any unused medication in the vial.
i. Blood will be drawn for trough drug concentrations at visits 2, 4, 6, 8, 9, 10, and 11 (weeks 2, 4, 8, 12, 16, 20, 24 after commencing medication).
j. Blood and urine will also be collected at visits 2, 4, 6, 8, 9, 10, and 11 for drug screen, complete biochemistry and hematology analysis.
k. Subjects will be referred to their family physicians either at the end of the 6 months or if a subject decides to terminate participation in the study.
l. Individuals who do not respond to drug XXX will be referred to alternate psychiatric treatment or to their family physicians.

**Follow-up visits 11 and 12 (at 8 months and at the 12 months after treatment)**

a. Review daily diary.
b. Medical examination.
c. Psychiatrist: examine any increased anxious symptoms, interview patients for concomitant illness, and examine potential adverse reactions.
d. Complete questionnaires: HAM-A, STAI-XI, VAS.
e. Blood and urine collection for drug screen, complete biochemistry, and hematology analyses.

**II.2.f. Planned Sample**

Refer to Flemming’s Single Stage Procedure Subsection V (planned sample) of the Mood disorders Section II.2.e. in Chapter 20.
II.2.f. Study Population
Male or Female over 18 years of age meeting DSM-IV criteria for GAD and who exhibit moderate to severe anxiety or a Ham-A score of $\geq 18$, and a score of $\geq 2$ on the HAM-A item 1 (anxious mood) and item 2 (tension).

II.2.g. Specific Inclusion Criteria
A subject will be eligible for inclusion in the study only if all of the following criteria apply:

- a. Males or females between 19 to 50 years of age.
- b. Socially stable.
- c. Meet DSM-IV criteria for general anxiety disease.
- d. In-patients or out-patients.
- e. Non smokers.

II.2.h. Specific Exclusion Criteria
Exclusion criteria must take into account the characteristics of the drug (pharmacokinetics, and pharmacodynamics, drug-drug interactions, and adverse effects). Patients will not be eligible to participate in the study if any of the following criteria apply:

- a. If they meet criteria for comorbid anxiety and MDD, or Panic disorder, or Obsessive compulsive disorder.
- b. Active suicidal ideation.
- c. If they meet criteria for Bipolar Disorder, schizophrenia, schizo-affective or other substance abuse/dependence.
- d. Evidence of medical or surgical illness requiring treatment.
- e. History of psychoactive drug dependence or a positive urine test for psychoactive drugs
- f. Use of medications which may interfere with the study procedures (e.g. SSRIs).
- g. Any clinically significant abnormality evident in biochemistry or hematology test results or in urine analysis requiring further investigation.
- h. Receiving or will receive other investigational drug during the study.
- i. Pregnant or lactating females.

II.2.i. Tools to Assess Endpoints
The tools used to assess efficacy are shown in the following table.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Assessment Tools to Measure Variable</th>
<th>Time of Assessment</th>
</tr>
</thead>
<tbody>
<tr>
<td>Diagnosis of Anxiety</td>
<td>DSM-IV, HAM-A</td>
<td>Visit 1 (baseline)</td>
</tr>
<tr>
<td>Level and Severity of Anxiety</td>
<td>HAM-A, VAAS</td>
<td>Visit 1 (baseline)</td>
</tr>
<tr>
<td>Reduction in symptoms of anxiety</td>
<td>HAM-A, Sta-IX, VAAS</td>
<td>Visits 1-12</td>
</tr>
<tr>
<td>Reduction in Trait Effects</td>
<td>Sta-IX</td>
<td>Visits 1-12</td>
</tr>
<tr>
<td>Reduction in Treatment Effects</td>
<td>Sta-IX</td>
<td>Visits 1-12</td>
</tr>
<tr>
<td>Time of onset of a consistent decrease in anxiety symptoms</td>
<td>HAM-A, VAAS</td>
<td>Visits 1-12</td>
</tr>
<tr>
<td>Time of greatest reduction in symptoms of anxiety</td>
<td>HAM-A</td>
<td>Visits 1-10</td>
</tr>
<tr>
<td>Number of patients achieving HAM-A &lt; 10</td>
<td>HAM-A</td>
<td>Visits 1-10</td>
</tr>
<tr>
<td>Drug Compliance</td>
<td>Patient Daily Diaries and Returned Medication</td>
<td>Visits 1-8</td>
</tr>
</tbody>
</table>
The tools to assess safety are described in Subsection II.2.i. (Tools to Assess Safety) of the Mood Disorders Section of Chapter 20.

II.2.j. Specific criteria for early withdrawal and discontinuation
Refer to Subsection II.2.j. (Specific criteria for early withdrawal and discontinuation) of the Mood Disorders Section of Chapter 20.

II.2.k. Data Analysis Method
Refer to Subsection II.2.k. (Data Analysis Method) of the Mood Disorders Section of Chapter 20.

III. REFERENCES

Chapter 23. Schizophrenic Disorders

Claudio A. Naranjo, M.D. 1

Lara Chayab 2

1 Professor
Department of Pharmacology
Psychiatry and Medicine
University of Toronto
Toronto, Ontario
CANADA
and
Head
Neuropsychopharmacology Research Program
Sunnybrook & Women’s College Health Science Centre
Toronto, Ontario
CANADA

2 M.Sc. Candidate
Department of Pharmacology
University of Toronto
Toronto, Ontario
CANADA
and
Researcher
Neuropsychopharmacology Research Program
Sunnybrook & Women’s College Health Science Centre
Toronto, Ontario
CANADA
I. INTRODUCTORY REMARKS

Psychotic disorders are personality and thought disorders that are associated with emotional and behavioral impairments (1). Schizophrenia is a psychotic disorder that is accompanied by impairments in speech patterns, inability to process information, delusions, and hallucinations (1). Schizophrenia consists of positive and negative symptoms (1). Increased activity, agitation, delusions, and hallucinations characterize positive symptoms (1). The mechanism by which these symptoms are manifest is through an increase in dopaminergic activity in the brain. Negative symptoms, on the other hand, are characterized by a decrease in activity, loss of pleasure, withdrawal from social interactions, a decrease in dopaminergic activity in the cerebral cortex, and an increase in dopaminergic activity in the stratum (1). The diagnostic criteria of Schizophrenia according to the DSM-IV include that the patient’s illness has to be continuous for at least six months and an occurrence of at least one psychotic phase followed by a residual phase (2). Pharmacological treatments of Schizophrenia include the use of first generation antipsychotics, such as haloperidol, and second generation antipsychotics, such as clozapine (3). First generation antipsychotics are potent antagonists of the dopamine 2 receptors in the stratum. They also block serotonergic, cholinergic, adrenergic, and histaminergic receptors (3). The wide array of affinity of these drugs to various receptors leads to side effects including extrapyramidal effects (sedation), tachycardia, dry mouth, blurred vision, gastrointestinal problems, sexual dysfunction, and orthostatic hypotension, Parkinson’s like syndrome, tardive dyskinesia, dystonia, akathisia, and neuroleptic malignant syndrome (1). These side effects are dose-related unlike the efficacy, which reaches a plateau after a certain dose. All first generation (typical) antipsychotics exhibit the same efficacy for the treatment of positive symptoms of schizophrenia, but their effect on negative symptoms has yet to be determined. Second generation (atypical) antipsychotics were developed for patients who are resistant to treatment with first generation antipsychotics (1). They have been shown in many studies to exhibit higher efficacy and slightly better tolerability (better safety profile) than first generation drugs in the treatment of schizophrenia (1). Many atypical antipsychotics block the dopamine 4 receptors in the cerebral cortex. They also have affinity for the serotonin 2 receptors and the dopamine 2 receptors (1). The better tolerability that atypical drugs exhibit accounts for the lower risk of extrapyramidal symptoms; however, adverse effects on the cholinergic and adrenergic systems still exist (1). As a result, an antipsychotic drug is required that exhibits high efficacy for the treatment of both positive and negative symptoms of schizophrenia. In addition, the antipsychotic drug must exhibit a wide therapeutic index, no extrapyramidal side effects, and very low cholinergic and adrenergic effects.

In this chapter, a double blind, placebo controlled, parallel-group study design will be used to test the efficacy and safety of investigational drug YYY versus placebo in patients diagnosed with schizophrenia. Previous clinical studies have shown that drug YYY is a highly potent dopamine 4 receptor antagonist, has no sedative effect, has a fast onset of action, and requires once daily dosing. In this study design, these properties of the drug will be better characterized and its efficacy and safety will be determined. The example of the study design illustrated in this chapter is most applicable to trials evaluating antipsychotic drugs intended to be used in patients with schizophrenia.

II. PHASE II STUDIES FOR REGISTRATION OF NEW ANTIPSYCHOTIC DRUGS

II.1. Outline of a typical development plan

This study will examine the efficacy and safety of drug YYY in men and women diagnosed with schizophrenia according to the DSM-IV. All patients enrolled in the study meeting inclusion/exclusion criteria and that give consent will be randomly assigned to receive one oral daily dose of drug YYY or placebo for 6 months. A long-term trial has been chosen for this study since the cholinergic and adrenergic side effects usually take 3 to 6 months to appear. Efficacy and safety measures are going to be performed
at weekly intervals up to the end of week 12, and then at 2-week intervals up to week 24. The study will be randomized, double blind, placebo-controlled and patients will come back for two follow-up assessments.

II.2. Long-term studies

II.2.a. Study Objectives

Primary Objectives
a. To compare the efficacy of drug YYY treatment versus placebo in reducing the positive and negative symptoms of schizophrenia.
b. To compare the safety of drug YYY treatment versus placebo

Secondary Objectives
a. To determine the onset of the antipsychotic action of drug YYY
b. To determine the duration of the antipsychotic effect of drug YYY
c. To determine the peak antipsychotic effects of drug YYY

II.2.b. Primary Endpoints

Rating scales will be administered to assess the following dependent variables:
a. Structured Clinical Interview for DSM-IV to diagnose patients with schizophrenia.
b. Brief Psychiatric Rating Scale (BSRS) to determine the severity of the disorder and positive and negative symptoms. It is a 16-item scale, which includes 7 points for severity scale, 5 points for positive symptoms, 2 points for negative symptoms, and 9 general symptom points. Patients with schizophrenia score $\geq 33$ points out of 112 (4). Positive and Negative Syndrome Scale (PANSS) which also determines the severity of schizophrenia and more specifically deals with the positive and negative symptoms associated with schizophrenia. It is a 30-item scale, which includes 7 points that measure positive symptoms, 7 points for negative symptoms, and 16 points for general psychopathology symptom measure. A schizophrenic patient would typically score 91 at beginning of trial (4).
c. Clinical Global Impression Scale (CGI) is used to assess treatment response in psychiatric patients. It is a 3-item scale that measures the severity of the illness (7-point scale), global impairment (7-point scale), and efficacy index (4-point scale). This scale is taken at baseline and repeated after drug exposure in order to compare results and assess efficacy (4).
d. Simpson Angus Scale (SAS) detects any drug induced parkinsonism and extrapyramidal side effects. It evaluates the presence and severity of the symptoms using a 10-item rating scale (4).
e. Barnes Akathisia Scale (BAS) is a 4-item scale that detects the presence and severity of any drug induced akathisia. The scale measures the objective and subjective effects such as restlessness and awareness of restlessness respectively (4).
f. Extrapyramidal Symptom Rating Scale (ESRS) is a 12-item scale that detects the presence of drug induced parkinsonism like symptoms, akathisia, dyskinesia, and dystonia (4).
g. Abnormal Voluntary Movement Scale (AIMS) is a scale that detects the patient’s movements by providing certain positions in which the patients have to rotate his/her body and the psychiatrists assess whether abnormal facial or body movements exist (4).

Responders to the antipsychotic drug YYY versus placebo, where a response is defined as:
a. $R =$ a reduction from baseline (visit 1) during (weeks 1 – 24) and post treatment (2 follow-up sessions) as measured by the BSRS.
b. $R =$ a reduction from baseline (visit 1) during (weeks 1 – 24) and post treatment (2 follow-up sessions) as measured by the PANSS.
c. \( R \) = a reduction in the CGI scale scores on weeks 1-24 compared to baseline will indicate the presence of drug efficacy. The higher a reduction in the score, the more efficacious the drug is considered.

d. \( R \) = an increase in the SAS scale scores on weeks 1-24 compared to baseline will indicate a presence of parkinsonism-like adverse effect.

e. \( R \) = an increase in the BAS scale scores on weeks 1-24 compared to baseline will indicate the presence of akathasia.

f. \( R \) = an increase in the ESRS scale scores on weeks 1-24 compared to baseline will indicate the presence of extrapyramidal adverse effects.

g. \( R \) = an increase in the AIMS scale score on weeks 1-24 compared to baseline will indicate the presence of abnormal voluntary movements.

II.2.c. Secondary Endpoints

a. The time of onset of a consistent decrease in schizophrenic symptoms as measured by the CGI compared to baseline.

b. The treatment day during which the greatest reduction in the schizophrenic symptoms is present as measured by the PANSS.

c. The number of patients that achieve a BSRS < 33 and a PANSS < 91 by the end of the study period.

d. Time of onset of any adverse effect as measured by the SAS, BAS, ESRS, and AIMS scales.

e. Severity of the adverse effects (if developed) as measured by the SAS, BAS, ESRS, and AIMS scales.

f. Number of subjects that drop out because of the development of adverse effects.

Adverse effects are measured by changes in blood pressure, heart rate, GI motility, and abnormal laboratory tests (blood and liver).

II.2.d. Study Design

This is a double blind, randomized, placebo-controlled, parallel-group study. There will be 2 groups in this study, schizophrenic patients and a healthy control group. Patients in each group will receive either a single dose of drug YYY or a single dose of placebo randomly once daily for 24 weeks. The number of patients receiving drug YYY will equal the number of patients receiving placebo within each group. The antipsychotic effect of drug YYY will be measured using various questionnaires (discussed below). Objective measures, such as blood pressure, will also be obtained. The results will be compared and analyzed between and within groups.

Screening for Eligibility (Visit 1, One week before start of the study)

Subjects will be interviewed to ensure suitability for study participation a week before the commencement of the study. The following will be required to assess eligibility:

a. Written informed consent.

b. Structured Clinical Interview for DSM-IV to assess psychiatric status and to rule out dependence on psychoactive substances.

c. Current level of schizophrenia (Brief Psychiatric Rating Scale (BSRS), Positive and Negative Syndrome Scale (PANSS), Clinical Global Impression Scale (CGI), Simpson Angus Scale (SAS), Barnes Akathisia Scale (BAS), Extrapyramidal Symptom Rating Scale (ESRS), Abnormal Voluntary Movement Scale (AIMS)). This is the baseline measure to which all upcoming results will be compared.

d. Brief medical examination (heart rate, blood pressure).

e. Medical history.

f. Review of inclusion/exclusion criteria.
g. Pregnancy test for women.
h. Blood and urine samples to assess liver function, hematology, biochemistry, and to detect the presence of other psychoactive drugs.
i. Participants will be provided with a pager number to be used if they experience any serious side effects and a wallet card containing information about participation in this study.
j. Patients are also given a diary in which they record drug compliance and any side effects that they may experience on a daily basis.

Treatment Phase (Visits 2-19, Months 1-6)

a. Eligible subjects will attend 24 treatment sessions (weekly intervals up to the end of the 6 months).
b. Medical examination.
c. Medication will be dispensed (enough pills to last for the next visit at once daily dosing)
d. Treatment will consist of 60 minutes sessions with the research assistant.
e. A psychiatrist will be available for consultation, assessment, and treatment as needed (i.e. adverse drug reaction, increases in severity of anxious symptomatology).
f. Review Daily Diary forms on which patients record compliance with medication and any side effects that the patient may be experiencing.
g. At each visit, the BSRS, PANSS, SAS, BAS, ESRS, and AIMS will be completed and subjects will be interviewed regarding concomitant illness and medication use.
h. Ask patients to return any unused medication in the vial.
i. Blood will be drawn for trough drug concentrations at 2-week intervals after the start of the study.
j. Blood and urine will also be collected at 2-week intervals for drug screen, complete biochemistry and hematology analysis.
k. Subjects will be referred to their family physicians either at the end of the 6 months or if a subject decides to terminate participation in the study.
l. Individuals who do not respond to drug YYY will be referred to alternate psychiatric treatment or to their family physicians.

Follow-up visits (Visits 20-21, at 8 months and 12 months after treatment)

a. Review daily diary.
b. Medical examination.
c. Psychiatrist: examine any increased schizophrenic symptoms, interview patients for concomitant illness, and examine potential adverse reactions.
d. Complete questionnaires: BSRS, PANSS, SAS, BAS, ESRS and AIMS.
e. Blood and urine collection for drug screen, complete biochemistry and hematology analysis.

II.2.e. Planned Sample
Refer to Flemming’s Single Stage Procedure Subsection II.2.e. (Planned sample) of the Mood disorders Section in Chapter 20.

II.2.f. Study Population
Male or female over 18 years of age meeting DSM-IV criteria for Schizophrenia and score ≥33 on the BSRS and ≥ 91 on the PANSS.

II.2.g. Specific Inclusion Criteria
A subject will be eligible for inclusion in the study only if all of the following criteria apply:
   a. Males or females between 19 to 50 years of age.
   b. Socially stable.
c. Meet DSM-IV criteria for schizophrenia.
d. In-patients or out-patients.
e. Non-smokers.

II.2.h. Specific Exclusion Criteria
Exclusion criteria must take into account the characteristics of the drug (pharmacokinetics, and pharmacodynamics, drug-drug interactions, and adverse effects). Patients will not be eligible to participate in the study if any of the following criteria apply:
a. If meet criteria for Bipolar Disorder, MDD, Anxiety, or substance abuse/dependence.
b. Evidence of medical or surgical illness requiring treatment.
c. History of psychoactive drug dependence or a positive urine test for psychoactive drugs.
d. Use of medications which may interfere with the study procedures (e.g. SSRIs).
e. Any clinically significant abnormality evident in biochemistry or hematology test results or in urine analysis requiring further investigation.
f. Receiving or will receive other investigational drug during the study.
g. Pregnant or lactating females.

II.2.i. Tools to assess endpoints
The tools used to assess efficacy are shown in the following table.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Assessment Tools to Measure Variable</th>
<th>Time of Assessment</th>
</tr>
</thead>
<tbody>
<tr>
<td>Diagnosis of Schizophrenia</td>
<td>DSM-IV</td>
<td>Visit 1 (baseline)</td>
</tr>
<tr>
<td>Level and Severity of Schizophrenia</td>
<td>BSRS, PANSS, CGI</td>
<td>Visit 1 (baseline)</td>
</tr>
<tr>
<td>Reduction in Schizophrenic Symptoms</td>
<td>BSRS, PANSS, CGI</td>
<td>Visits 2-21</td>
</tr>
<tr>
<td>Reduction in Positive Symptoms</td>
<td>BSRS, PANSS</td>
<td>Visits 2-21</td>
</tr>
<tr>
<td>Reduction in Negative Symptoms</td>
<td>BSRS, PANSS</td>
<td>Visits 2-21</td>
</tr>
<tr>
<td>Induction of Extrapyramidal Adverse Effects</td>
<td>SAS, ESRS</td>
<td>Visits 2-21</td>
</tr>
<tr>
<td>Time of induction and Severity of Extrapyramidal AE</td>
<td>SAS, ESRS</td>
<td>Visits 2-21</td>
</tr>
<tr>
<td>Induction of Parkinsonism-like Symptoms</td>
<td>SAS</td>
<td>Visits 2-21</td>
</tr>
<tr>
<td>Time of Induction and Severity of Parkinsonism-like Symptoms</td>
<td>SAS</td>
<td>Visits 2-21</td>
</tr>
<tr>
<td>Induction of Akathisia</td>
<td>BAS</td>
<td>Visits 2-21</td>
</tr>
<tr>
<td>Time of Induction and Severity of Akathisia</td>
<td>BAS</td>
<td>Visits 2-21</td>
</tr>
<tr>
<td>Induction of Involuntary Movements</td>
<td>AIMS</td>
<td>Visits 2-21</td>
</tr>
<tr>
<td>Time of Induction and Severity of Involuntary Movements</td>
<td>AIMS</td>
<td>Visits 2-21</td>
</tr>
<tr>
<td>Time of onset of a consistent decrease in schizophrenic symptoms</td>
<td>BSRS, PANSS, CGI</td>
<td>Visits 2-21</td>
</tr>
<tr>
<td>Time of greatest reduction of Schizophrenic symptoms</td>
<td>BSRS, PANSS, CGI</td>
<td>Visits 2-21</td>
</tr>
<tr>
<td>Number of patients achieving BSRS &gt; 33</td>
<td>BSRS</td>
<td>Visits 2-21</td>
</tr>
<tr>
<td>Number of patients achieving PANSS &gt; 91</td>
<td>PANSS</td>
<td>Visits 2-21</td>
</tr>
<tr>
<td>Drug Compliance</td>
<td>Patient Daily Diaries and Returned Medication</td>
<td>Visits 2-19</td>
</tr>
</tbody>
</table>
The tools to assess safety are described in Subsection II.2.i. (Tools to Assess Safety) of the Mood Disorders Section of Chapter 20.

**II.2.j. Specific criteria for early withdrawal and discontinuation**
Refer to Subsection II.2.j. (Specific criteria for early withdrawal and discontinuation) of the Mood Disorders Section of Chapter 20.

**II.2.k. Data Analysis Method**
Refer to Subsection II.2.k. (Data Analysis Method) of the Mood Disorders Section of Chapter 20.

**III. REFERENCES**

Chapter 24. Alcohol and Nicotine Addiction

Claudio A. Naranjo, M.D. ¹

Lara Chayab ²

¹ Professor
Department of Pharmacology
Psychiatry and Medicine
University of Toronto
Toronto, Ontario
CANADA
and
Head
Neuropsychopharmacology Research Program
Sunnybrook & Women’s College Health Science Centre
Toronto, Ontario
CANADA

² M.Sc. Candidate
Department of Pharmacology
University of Toronto
Toronto, Ontario
CANADA
and
Researcher
Neuropsychopharmacology Research Program
Sunnybrook & Women’s College Health Science Centre
Toronto, Ontario
CANADA
I. INTRODUCTORY REMARKS

Alcohol, opioids, nicotine, and psychostimulants have different chemical structures; however, they seem to exert their actions via similar neurochemical pathways in the brain, which lead to addiction. Current pharmacotherapy approaches available for the treatment of alcohol and drug addiction aim at minimizing symptoms of acute abstinence and the risk of relapse. Alcoholism is a complex disorder exhibiting multiple symptoms. It is often co-morbid with Major Depressive Disorders, antisocial personality, or anxiety. According to the American Psychiatric Association (DSM-IV), individuals must meet three of the following criteria during a 12-month period for a diagnosis of alcohol dependence: a) Tolerance to alcohol, increase amounts of alcohol consumption to achieve same effects, b) Signs or symptoms of alcohol withdrawal, c) Attempts to cut down are unsuccessful, d) Long periods of time spent in obtaining alcohol, alcohol consumption, and hangovers, e) Impaired social and work activities due to alcohol consumption, and f) Alcohol consumption is not decreased even if it leads to adverse effects physically and psychologically (1).

Pharmacological treatment of alcohol dependence include agents that minimize the positive reinforcing effects of alcohol (such as naltrexone) and other agents used to relieve withdrawal symptoms and promote abstinence (such as Sedatives and Disulfiram) (2). Naltrexone is a mu-opioid and a delta-opioid receptor antagonist. It functions by blocking the binding of the endogenous opioid, beta-endorphin, to the mu-opioid receptor, which leads to the alleviation of positive effects (euphoria) induced by alcohol intake (3). It is administered orally, and taken 3 times/week at 100-150mg. It is generally safe, with no known interactions caused by alcohol intake and no withdrawal symptoms after drug discontinuation (3). Studies have shown that naltrexone does not lead to complete abstinence from drinking, however, it may cause a reduction in the amount of alcohol intake and a better control over drinking behaviors (4). Acamprosate exhibits a structure similar to the neurotransmitters GABA and glutamate (5). It is thought to work by stabilizing the neurotransmitter balance seen in alcohol dependent people (5). Several trials have shown that acamprosate is efficacious in the treatment of alcohol dependence and is well tolerated (5). However, approval of acamprasate is still pending in many countries. As a result, there is a need for pharmacological agents that treat alcohol dependence, maintain abstinence from alcohol, and do not exhibit adverse effects.

Persistence of cigarette smoking leads to nicotine addiction. Smoking is the leading cause of death in North America, implicated in one of every five deaths (4). Unaided attempts of smoking cessation are successful in only 5% of people who attempt to quit (4). Most pharmacological agents that are available for smoking cessation are nicotine replacement agents such as nicotine (a chewing gum formulation that contains 2 mg of nicotine), and nicotine patch (6). In addition, some investigational drugs for smoking cessation include nicotine inhalers, mecaminamine, a nicotine receptor antagonist, antidepressants, clonidine, and airway sensory replacement (6). Nicorette produces adverse effects that include bad taste, difficulty with chewing, and stomach upset (6). The nicotine patch is better tolerated; however, it may lead to skin irritation and allergies in patients (6). As a result, a pharmacological agent is required that functions to cease cigarette smoking in individuals that are mild to heavy smokers, decreases craving, has no side effects and prevents relapse.

In this chapter, two separate double blind, placebo-controlled, parallel-group study designs will be carried out to test the efficacy and safety of investigational drug AAA (in study-1 for patients with moderate to severe depression, and investigational drug SSS (in study-2 for patients with nicotine addiction. Previous clinical studies have shown that drug AAA is a highly purely selective mu-opioid receptor antagonist, has a half-life of 24 hrs, and is administered 3 times/week. In addition, drug AAA has no potential for alcohol interactions, has mild adverse effects, and has no potential for withdrawal symptoms after drug discontinuation. With the design of the present study, the properties of drug AAA will be better characterized and its efficacy will be determined. Drug SSS is a highly selective nicotine receptor
antagonist that blocks the physiological, behavioral, and reinforcing effects of nicotine. It has a half-life of 12 hours, which accounts for its twice daily dosing, and it is well tolerated due to its mild adverse effects. The properties of Drug SSS will be further investigated and its efficacy will be determined with the present experimental design. The experimental design of both study is most applicable to trials intended to be used in patients as a first line of treatment.

II. PHASE II STUDIES FOR REGISTRATION OF NEW DRUGS FOR THE TREATMENT OF ALCOHOL AND NICOTINE DEPENDENCE

II.1 Outline of a typical development plan

The two studies will examine the efficacy and safety of drug AAA in men and women experiencing moderate to severe alcohol dependence and drug SSS in men and women experiencing moderate to severe nicotine addiction. All patients enrolled in study 1 meeting inclusion/exclusion criteria for alcohol dependence and that give consent will be randomly assigned to receive one oral dose of drug AAA 3 times/week or placebo for 3 months. In addition, patients enrolled in study 2 meeting inclusion/exclusion criteria for nicotine addiction will be randomly assigned two oral daily doses (every 12 hours) of drug SSS or placebo for 3 months. Efficacy and safety measures are going to be performed at weekly intervals up to the end of the third month. The study will be a randomized, double blind, placebo-controlled and patients will come back for two follow-up assessments.

II.2. Short-term studies

II.2.a. Study Objectives

Primary objectives of study 1
a. To compare the efficacy of drug AAA treatment versus placebo in the abstinence from alcohol in alcohol dependent patients.

b. To compare the safety of drug AAA treatment versus placebo.

Secondary objectives for study 1
a. To determine the onset of a decrease in alcohol intake by drug AAA.
b. To determine the duration of the decrease in alcohol intake by drug AAA.
c. To determine the time at which a complete abstinence from alcohol takes place.
d. To determine the time at which a consistent abstinence from alcohol takes place.
e. To determine if patients will relapse after abstinence.
f. The time at which patients start drinking again if relapse took place.

Primary objectives of study 2
a. To compare the efficacy of drug SSS treatment versus placebo in the abstinence from cigarette smoking in patients with nicotine addiction.
b. To compare the safety of drug SSS treatment versus placebo.

Secondary objectives for study 2
a. To determine the onset of a decrease in cigarette smoking by drug AAA.
b. To determine the duration of the decrease in cigarette smoking by drug AAA.
c. To determine the time at which a complete abstinence from cigarette smoking takes place.
d. To determine the time at which a consistent abstinence from cigarette smoking takes place.
e. To determine if patients will relapse after abstinence.
f. The time at which patients start drinking again if relapse took place.
II.2.b. Primary Endpoints

Primary endpoints of study 1

Rating scales will be administered to assess the following dependent variables:

- a. Structured Clinical Interview for DSM-IV to diagnose patients with alcohol dependence.
- b. Alcohol Dependence Scale (ADS): The ADS provides a quantitative measure of the severity of alcohol dependence consistent with the concept of the alcohol dependence syndrome. Its 25 items cover alcohol withdrawal symptoms, impaired control over drinking, awareness of a compulsion to drink, increased tolerance to alcohol, and salience of drink-seeking behaviour. Alcohol dependent patients entering the study must score ≥ 22 on the ADS, while healthy individuals must score ≤ 2 (7).
- c. Michigan Alcohol Screening Test (MAST): Consisting of 25 questions, the MAST serves to uncover the problems the individual is experiencing as a result of his/her alcohol dependence. Because of the seemingly neutrality of some of the questions, it is easier to extract pertinent information about one's affliction which that person might have been otherwise reluctant to admit. Alcoholic dependent patients entering the study must score ≥ 6 while healthy patients in the control group must score ≤ 2 (7).
- d. Clinical Institute Withdrawal Assessment for Alcohol (revised) (CIWA): This 10-item scale is used to measure the severity of alcohol withdrawal symptoms. It is important for our results that subjects are not undergoing withdrawal while being tested. All patients entering the study must score ≤ 15 for no signs of withdrawal symptoms (7).

Responders to drug AAA versus placebo, where a response is defined as:

- a. R = a reduction from baseline (visit 1) on weeks 1-12 during and post treatment as measured by the ADS.
- b. R = a score of ≤ 2 as measured by the MAST during (weeks 1-12) and post treatment (follow-up sessions) with drug AAA.
- c. R ≤ 15 on the CIWA for no signs of withdrawal detected.

Primary endpoints of study 2

Rating scales will be administered to assess the following dependent variables:

- a. Structured Clinical Interview for DSM-IV to diagnose patients with nicotine dependence.
- b. Fagerstrom Test for Nicotine Dependence (FTND) is used to assess tobacco dependence. The questionnaire contains items that determine the number of cigarettes smoked per day, the time to the first cigarette after awakening, and the difficulty of restraining from smoking when strongly advised to (ill). Patients with moderate to severe nicotine dependence typically score ≥ 6. Non-smokers (in the control group) should have 0 points (7).

Responders to drug SSS versus placebo, where a response is defined as:

- a. R = any reduction from baseline (visit 1) on weeks 1-12 during and post treatment as measured by the FTND.

II.2.c. Secondary endpoints

Secondary endpoints for study 1

- a. The first day during which a reduction in alcohol intake is seen as measured by ADS and as reported by the patient daily diaries.
- b. The treatment day during which the greatest reduction in alcohol intake is present as measured by the ADS.
- c. The day at which a complete abstinence from alcohol takes place as measured by the ADS and as reported by the daily diaries.
- d. The amount of days during which a consistent abstinence from alcohol takes place.
e. The number of patients that achieve ADS ≤ 2.
f. The day at which patients start drinking alcohol again if relapse occurs.

Secondary endpoints for study 2
a. The first day during which a reduction in cigarette smoking is seen as measured by FTND and as reported by the patient daily diaries.
b. The treatment day during which the greatest reduction in cigarette smoking is present as measured by the FTND.
c. The day at which a complete abstinence from cigarette smoking takes place as measured by the FTND and as reported by the daily diaries.
d. The amount of days during which a consistent abstinence from cigarette smoking takes place.
e. The number of patients that achieve FTND = 0.
f. The day at which patients start smoking again if relapse occurs.

Adverse effects are measured by changes in blood pressure, heart rate, GI motility, and abnormal laboratory tests (blood and liver).

II.2.d. Study Design
Both studies are double blind, randomized, placebo-controlled, parallel-group study. There will be 2 groups in study 1, a moderate to severely alcohol dependent group, and a healthy control group. Patients in each group will receive either a single dose of drug AAA or a single dose of placebo randomly once daily, 3 times/week, for 12 weeks. In addition, there will also be 2 groups in study 2, a moderate to severely nicotine dependent group, and a healthy (non-smokers) control group. Patients in each group will receive either drug SSS or placebo twice a day for 12 weeks. The number of patients receiving drug AAA/SSS will equal the number of patients receiving placebo within each group. The effects of drug AAA/SSS will be measured using various questionnaires (discussed below). Objective measures, such as blood pressure, will also be obtained. The results will be compared and analyzed between and within groups.

Screening for Eligibility (Visit 1)
Subjects will be interviewed to ensure suitability for study participation a week before the commencement of the study. The following will be required to assess eligibility:

a. Written informed consent.
b. Structured Clinical Interview for DSM-IV to assess dependence on psychoactive substances (alcohol or nicotine).
c. Current level of alcohol dependence (moderate to severe alcohol dependent patients must have a score of ≥ 22 on the ADS while healthy patients in the control group must score ADS ≤ 2). Current level of nicotine dependence (moderate to severe nicotine dependent patients must score ≥ 3 on the FTND while non-smokers in the control group must have no FTND score. This is the baseline measure to which all upcoming results will be compared against.
d. Brief medical examination (heart rate, blood pressure).
e. Medical history.
f. Review of inclusion/exclusion criteria.
g. Pregnancy test for women.
h. Blood and urine samples to assess liver function, hematology, biochemistry, and to detect the presence of other psychoactive drugs.
i. Participants will be provided with a pager number to be used if they experience any serious side effects and a wallet card containing information about participation in this study.
j. Patients are also given a diary in which they record drug compliance and any side effects that they may experience on a daily basis.
Treatment Phase (Visits 2-7, Weeks 1-12)

a. Eligible subjects will attend six treatment sessions (one every two weeks).
b. Medical examination.
c. Medication will be dispensed (enough pills for two weeks).
d. Treatment will take place at 2-week intervals consisting of 30 to 45 minute sessions with the research assistant.
e. A psychiatrist will be available for consultation, assessment, and treatment, as needed (i.e. adverse drug reaction, any withdrawal symptoms).
f. Review Daily Diary forms on which patients record compliance with medication.
g. Study 1: At each visit, the ADS, MAST, and CIWA will be completed and subjects will be interviewed regarding concomitant illness and medication use.
h. Study 2: At each visit, the FTND will be completed and subjects will be interviewed regarding concomitant illness and medication use.
i. Ask patients to return any unused medication in the vial.
j. Blood will be drawn for trough drug concentrations at visits 3, 5 and 7 (4, 8 and 12 weeks after commencing medication).
k. Blood and urine will also be collected at visits 4, 6 and 8 for drug screen, complete biochemistry and hematology analyses.
l. Subjects will be referred to their family physicians either at the end of the 12 week study or if a subject decides to terminate participation in the study.
m. Individuals who do not respond to drug AAA/SSS will be referred to alternate psychiatric treatment or to their family physicians.

Follow-up visits (Visits 8-9, at 3 months and 6 months after treatment)

a. Review daily diary.
b. Medical examination.
c. Psychiatrist: examine any increased alcohol/nicotine dependence, interview patients for concomitant illness, and examine potential adverse reactions.
d. Study 1: Complete questionnaires: ADS, MAST, and CIWA to check for increase/decrease/relapse to alcohol dependence.
e. Study 2: Complete questionnaires: FTND to check for increase/decrease/relapse to nicotine dependence.
f. Blood and urine collection for drug screen, complete biochemistry, and hematology analyses.

II.2.e. Planned Sample
Refer to Flemming’s Single Stage Procedure Subsection II.2.e. (planned sample) of the Mood disorders Section in Chapter 20.

II.2.f. Study Population
Males or Females over 18 years of age meeting DSM-IV criteria for alcohol/nicotine dependence and exhibit moderate to severe alcohol dependence or an ADS score of ≥ 22/ moderate to severe nicotine dependence or an FTND ≥ 3.

II.2.g. Specific Inclusion Criteria
A subject will be eligible for inclusion in both studies only if all of the following criteria apply:

a. Males or females between 19 to 50 years of age.
b. Socially stable.
c. Meet DSM-IV criteria for alcohol/nicotine dependence.
d. In-patients or out-patients.
e. Non-smokers for study 1.
II.2.h. Specific Exclusion Criteria
Exclusion criteria must take into account the characteristics of the drug (pharmacokinetics, and pharmacodynamics, drug-drug interactions, and adverse effects). Patients will not be eligible to participate in the study if one of the following criteria apply:

a. Meet criteria for MDD, Anxiety, Bipolar disorder, Schizophrenia, Schizo-affective or other substance abuse/dependence (other than alcohol and nicotine).

b. Evidence of medical or surgical illness requiring treatment.

c. History of psychoactive drug dependence (other than alcohol and nicotine) or a positive urine test for psychoactive drugs (other than alcohol and nicotine).

d. Use of medications which may interfere with the study procedures (e.g. SSRIs).

e. Any clinically significant abnormality evident in biochemistry or hematology test results or in urine analysis requiring further investigation.

f. Receiving or will receive other investigational drug during the study.

g. Pregnant or lactating females.

II.2.i. Tools to assess
Tools to assess efficacy in alcohol and nicotine dependence are shown in tables 1 and 2.

<table>
<thead>
<tr>
<th>Table 1. Alcohol Dependence</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Variable</strong></td>
</tr>
<tr>
<td>Diagnosis of Alcohol Dependence (AD)</td>
</tr>
<tr>
<td>Level and Severity of AD</td>
</tr>
<tr>
<td>Reduction in AD</td>
</tr>
<tr>
<td>Reduction in impaired control over drinking</td>
</tr>
<tr>
<td>Reduction in drinking-seeking behavior</td>
</tr>
<tr>
<td>Increase in tolerance to drinking</td>
</tr>
<tr>
<td>Severity of Alcohol Withdrawal</td>
</tr>
<tr>
<td>Time of onset of a consistent decrease in AD</td>
</tr>
<tr>
<td>Time of greatest reduction of AD</td>
</tr>
<tr>
<td>Number of patients achieving ADS &lt; 2</td>
</tr>
<tr>
<td>Number of patients achieving MAST &lt; 2</td>
</tr>
<tr>
<td>Drug Compliance</td>
</tr>
</tbody>
</table>

The tools to assess safety are described in Subsection II.2.i. (Tools to Assess Safety) of the Mood Disorders Section of Chapter 20.

II.2.j. Specific criteria for early withdrawal and discontinuation
Refer to Subsection II.2.j. (Specific criteria for early withdrawal and discontinuation) of the Mood Disorders Section of Chapter 20.

II.2.k. Data Analysis Method
Refer to Subsection II.2.k. (Data Analysis Method) of the Mood Disorders Section of Chapter 20.
Table 2. Nicotine Dependence

<table>
<thead>
<tr>
<th>Variable</th>
<th>Assessment Tools to Measure Variable</th>
<th>Time of Assessment</th>
</tr>
</thead>
<tbody>
<tr>
<td>Diagnosis of Nicotine Dependence (ND)</td>
<td>DSM-IV</td>
<td>Visit 1 (baseline)</td>
</tr>
<tr>
<td>Level and Severity of ND</td>
<td>FTND</td>
<td>Visit 1 (baseline)</td>
</tr>
<tr>
<td>Reduction in ND</td>
<td>FTND</td>
<td>Visits 2-9</td>
</tr>
<tr>
<td>Reduction in the number of cigarettes smoked per day</td>
<td>FTND</td>
<td>Visits 2-9</td>
</tr>
<tr>
<td>Reduction in the time to first cigarette after awakening</td>
<td>FTND</td>
<td>Visits 2-9</td>
</tr>
<tr>
<td>Time of onset of a consistent decrease in ND</td>
<td>FTND</td>
<td>Visits 2-9</td>
</tr>
<tr>
<td>Time of greatest reduction in ND</td>
<td>FTND</td>
<td>Visits 2-9</td>
</tr>
<tr>
<td>Number of Patients achieving FTND = 0</td>
<td>FTND</td>
<td>Visits 2-9</td>
</tr>
<tr>
<td>Drug Compliance</td>
<td>Patient’s Daily Diaries and Returned Medication</td>
<td>Visits 2-7</td>
</tr>
</tbody>
</table>

III. REFERENCES

Pharmacological Research in Joint Disorders

Chapter 25. Osteoarthritis/arthrosis Short Term Studies

Chapter 26. Osteoarthritis/arthrosis Long Term Studies

Chapter 27. Rheumatoid Arthritis
Chapter 25. Osteoarthritis/arthrosis Short Term Studies: Pain and Function Improvement

Jean-Pierre Raynauld, M.D. ¹

Johanne Martel-Pelletier, Ph.D. ¹

Boulos Haraoui, M.D. ¹

Jean-Pierre Pelletier, M.D. ¹

¹ Osteoarthritis Research Unit
Centre hospitalier de l’Université de Montréal (CHUM)
Hôpital Notre-Dame
Département de Medicine
Université de Montréal
Montréal, Québec
CANADA
I. INTRODUCTORY REMARKS

Osteoarthritis (OA) is the most common form of arthritis and is a leading cause of disability. More than 75% of those over age seventy exhibit radiographically detectable changes consistent with osteoarthritis. About 40-60% of subjects with radiological OA changes suffer from clinical symptoms such as pain, joint stiffness, and joint deformities.

Patients with OA have pain that typically worsens with weight bearing and activity and improves with rest, as well as morning stiffness, gelling of the involved joint after periods of inactivity, and limited joint motion. As OA progresses, pain at rest can also be present. With a few exceptions, the causes of OA are not known so that the main goals of therapy are pain relief and improved physical and social function.

Pharmacologic therapy of OA typically begins with analgesics such as acetaminophen in doses up to 4 g/day, progresses to low dose nonsteroidal anti-inflammatory drugs (NSAIDs), and then to full dose NSAIDs (including COX-2 selective inhibitors). NSAIDs, while useful, have a ceiling affect and can be limited in their use because of their side effects, particularly those affecting the gastrointestinal tract, liver, and kidney; the risks of which increase with advanced age.

The primary parameter of this study is the proportion of subjects who achieve adequate pain control (% with moderate, good, or excellent pain control) during 56 days of a new DRUG X. This parameter will be assessed with a pain control assessment performed at study entry, at each telephone contact, and at each visit. Among secondary parameters are the effect on pain intensity, quality of life, functionality, and global assessment of change.

II. RATIONALE FOR STUDY DESIGN

This trial is an observational, therapeutic use study investigating the effect of DRUG X treatment on pain control in subjects with moderate to severe pain due to OA of the hip or knee that is inadequately controlled with acetaminophen or a traditional NSAID.

DRUG X has been demonstrated to be safe and efficacious in chronic non-cancer pain in several randomised clinical trials. Moreover, DRUG X has been demonstrated to be preferred over several NSAIDs with the main reason being that better pain control was achieved. DRUG X was also associated with a better safety profile. The effect of 8 weeks treatment with DRUG X on pain control, quality of life, and functionality has not been previously investigated in a clinical study of subjects with moderate to severe OA pain of the hip or knee that is inadequately controlled with NSAIDs.

II.1. Outline of a Typical Development Plan

Multi-center, randomized, double-blind, placebo-controlled, parallel group study with mild to moderate primary knee (or hip) OA fulfilling the American College of Rheumatology (ACR) criteria (see Appendix A), who have been completely withdrawn from their previous analgesics or anti-inflammatory medications or have been newly diagnosed with mild to moderate primary knee (or hip) OA and who are not currently taking any analgesics or anti-inflammatory medications.

II.2. Short-Term Studies: Pain and Function Improvement

II.2.A. Objectives
The objective of this study is to investigate the short-term effect of a new DRUG X at a dose of Y mg compared to placebo, on pain control in subjects with moderate to severe pain due to OA according to the ACR (see Appendix A) of hip or knee that is inadequately controlled with simple analgesics or NSAIDs.
II.2.B. Primary endpoints
To determine the proportion of subjects with symptomatic osteoarthritis of the hip or knee who achieve adequate pain control (% of subjects with moderate, good, or excellent pain control) during 56 days of treatment with DRUG X.

II.2.C. Secondary endpoints
To compare the scores from the Numerical Pain Intensity Rating scale, WOMAC Osteoarthritis Index questionnaire and Acute SF-36 Health Survey after 56 days of treatment with DRUG X to baseline. Physician and Subject Global Impression of Change scales and Subject Treatment Assessment questionnaire will be done at the Final Visit.

II.2.D. Study design
This is a placebo-controlled, multicentre study.

Eligible subjects will undergo screening procedures. Subjects must show evidence of symptomatic hip or knee OA (ACR Functional Class ≥ grade 2) and meet the ACR hip or knee OA criteria, have “poor” or “very poor” pain control (on a five-point scale of excellent, good, moderate, poor, or very poor), and have at least moderate to severe pain demonstrated as a pain score ≥ 5 on a numerical pain intensity rating scale of 0 to 10 (with “0” representing no pain and “10” representing worst possible pain).

The duration of study treatment is 56 days. At Visit 1 (Study Entry) subjects will start on DRUG X at a dosage Y and will remain on this dose for the trial period.

During each study visit subjects will be required to complete a Pain Control Assessment indicating the amount of pain control experienced that day, a Western Ontario and McMaster Universities (WOMAC) Osteoarthritis Index and an SF-36 Acute Health Survey (Appendix B). Physician and Subject Global Impression of Change scales (7-point scale, see Efficacy Evaluations section) will be conducted at Visit 4. In the event that a subject withdraws early, all Visit 4 procedures will be performed.

At Visit 1 (Study Entry) subjects will start on DRUG X. Supplementary analgesic medication consisting of acetaminophen, 500 mg tablets, will be allowed during the study on an as needed basis, provided the total daily dose of acetaminophen does not exceed 4 g (8 tablets).

Eligible subjects who are enrolled into the study will be randomly allocated to be treated with DRUG X or a placebo for 56 days. Insufficient analgesia will be determined by the investigator, using his or her clinical judgement and taking into account the subject’s level of pain severity, level of pain control, use of supplementary acetaminophen, 500 mg tablets, and individual response and tolerance to the dose.

Subjects will be provided with acetaminophen, 500 mg tablets, as supplementary analgesic medication for any additional pain in the target OA hip or knee joint and will be taken as needed throughout the study (provided the total daily dose of acetaminophen does not exceed 4 g or 8 tablets). Subject use of acetaminophen, 500 mg tablets will be recorded in the Patient diary on a daily basis (Appendix C).

Concomitant analgesic opioid medication is NOT allowed during the course of the study. Weak opioid medication must be discontinued at the time of study entry.

Inhaled steroids for asthma or topical corticosteroid preparations for minor dermatological use will be allowed during the study.

Alcoholic beverages and sedating antihistamines may also produce additive depressant effects and should be used with caution.
ASA (acetylsalicylic acid) for cardiac prophylaxis, up to 325 mg/day, will be allowed during the study and should be used with caution with concomitant use of acetaminophen.

All medications (prescriptions or over-the-counter (OTC) medications, including supplements or nutraceuticals) and medical procedures ongoing in the week preceding study entry that are continued at the start of the study or are started during the trial and are different from the trial medication, must be documented on the Concomitant Therapy Form of the CRF. If any medication or medical procedure is started, stopped, or if the dose or frequency is modified, this must also be documented on the CRF. The sponsor must be notified in advance (or as soon as possible thereafter) of any instances in which prohibited therapies are administered.

For any concomitant therapy given as a treatment for a new condition or a worsening of an existing condition, the condition must be documented on the Adverse Event Form of the CRF.

III.2.F. Planned sample
Approximately 80 subjects will be required to be able to estimate the proportion of subjects achieving "excellent", "good" or "moderate" pain control within ± 11% with DRUG X compared to placebo with 95% confidence. Because only Visit 2 efficacy response is required for the evaluable population, no contingency is built in the required sample size.

Approximately 80 subjects who have a history of symptomatic OA of the hip or knee with chronic pain for at least 3 months, who have been on a stable daily dose of acetaminophen for at least two weeks prior to Study Entry, and who have uncontrolled pain, will be enrolled into the study.

Approximately 80 subjects from 10 sites will be screened and enrolled in the study. Each site will enrol approximately X subjects. Subjects are OA subjects, suffering with chronic hip or knee pain for at least 3 months (for at least 20 days of each month), and who are not hospitalised. The target joint selected will be an OA hip or knee joint that causes the most pain to the subject. In case of pain of equal severity in hip and knee, one target joint must be selected.

III.2.G. Study population
Potential study subjects must have a history of symptomatic osteoarthritis of the hip or knee with chronic pain for at least 3 months and must have been on a stable daily dose of acetaminophen at least 2 weeks prior to the study. The study will be explained to subjects and informed consent will be obtained.

III.2.H. Specific Inclusion Criteria
Subjects must satisfy the ALL of the following inclusion criteria before entering the study:

a. Male or female of ages ≥ 40.
b. Must be in generally good health as confirmed by medical and previous medication history, and baseline physical examination including vital signs.
c. Female subjects must be postmenopausal for at least 2 years, surgically sterile, or practising an effective method of birth control prior to entry and throughout the study, and have a negative urine pregnancy test at the baseline visit. The subject may continue in the study using abstinence as a form of birth control provided that she is completely abstinent, has a negative urine pregnancy test prior to study entry and at the final visit or upon termination (if the subject discontinues the trial early). It must be documented in the medical notes that the subject has been counselled about the birth control and the risks of becoming pregnant.
d. Symptomatic OA of the target hip or knee joint as evidenced by hip or knee pain for at least 3 months (for at least 20 days of each month) and osteophytes confirmed by an x-ray taken within the last two years and who must meet the OA hip or knee criteria of the American College of Rheumatology (Appendix A).
e. After a full explanation of the study, subjects must understand the nature of the study and sign the informed consent form to participate.

f. Subjects with moderate to severe pain of the target OA hip or knee joint whose pain is not adequately controlled with an NSAID. This will be defined as subjects with a pain control assessment of “poor” or “very poor” (on a five-point scale: excellent, good, moderate, poor, or very poor) and a mean pain score $\geq 5$ (on a numerical pain intensity rating scale of 0-10) at Baseline.

### III.2.1. Specific Exclusion Criteria

Potential subjects who meet any ONE of the following exclusion criteria will NOT be eligible to participate in the study:

a. Subjects who have previously failed on DRUG X therapy or those who previously have discontinued DRUG X due to adverse events.

b. Subjects who have received treatment with a strong opioid (e.g. morphine, hydromorphone, methadone, long-acting oxycodone, oxymorphone, levorphanol, heroin, etc.) in the 4 weeks preceding study entry. Subjects cannot take strong opioids during the study.

c. Subjects for whom a treatment is planned within the study period that could alter the degree or nature of pain (e.g. arthroscopic techniques, osteotomy, joint replacement surgery, etc.).

d. Subjects who are experiencing another type of continuous pain that is more severe in intensity in comparison with the OA target joint pain (e.g. low back pain, fibromyalgia, ankylosing spondylitis, etc.).

e. Subjects who have had target joint intra- or periarticular corticosteroid injections within 6 weeks of study entry or hyaluronic injections within 6 months of study entry. Injections are not allowed during the study. Subjects cannot have had arthrosynthesis within 4 weeks or arthroscopic techniques (e.g. joint débridement, abrasion, arthroplasty, chondral holes, etc.) within 3 months prior to the study or during the study.

f. Subjects taking glucosamine will not be eligible unless they have been on a stable dose for greater than 2 months preceding study entry. If subjects were taking a stable dose for at least 2 months prior to the study, the dosage should remain constant throughout the study. Glucosamine cannot be started at anytime during the study.

g. Subjects taking NSAIDs, COX-2 selective inhibitors, or steroidal drugs for at least 4 weeks before study entry may continue these medications during the study; however, they must have been taking a stable dose (consistent daily milligram dose $\pm 25\%$) for at least 2 weeks before study entry and the dosage must be kept constant throughout the study. If these medications were started within the 4 weeks preceding the study, the subject will be excluded, but can be rescreened at a later time. These medications cannot be started at anytime during the study.

h. Subjects who have had major surgery in the 3 months preceding the study.

i. Subjects with a significant psychiatric disorder (including major depression) or subjects receiving anti-psychotic medication.

j. Subjects who have taken sedatives, hypnotics, phenothiazines, anticonvulsants, tranquilizers or muscle relaxants two weeks preceding study entry. These medications cannot be started during the study.

k. Subjects who are taking tricyclic antidepressants if not expected to remain on a stable dose of these medications for the duration of the study. These medications cannot be started during the study.

l. Subjects who have applied topical analgesic preparations to the target joint and/or taken general anaesthetics in the one week preceding study entry. These medications cannot be started during the study.

m. Subjects with documented or suspected history of alcohol or drug abuse, or who have a documented or suspected history of an addictive personality.

n. Subjects who have started any form of physiotherapy, acupuncture, TENS, massage or active physical therapy within the 4 weeks preceding study entry. Such therapies can continue if they were started more than 4 weeks before the start of the study and if they continue at the same frequency of administration throughout the study. Any such therapies cannot be started during the study.
Female subjects who are breast-feeding.

Subjects known to have any of the following:

- significantly abnormal renal or hepatic function;
- any disease or condition that compromises the function of those body systems that could result in altered absorption, excess accumulation, or impaired metabolism or excretion of the test medications;
- a life-threatening disease (e.g. AIDS, malignant disease, etc.) that would preclude completion of study or interfere with protocol compliance;
- any condition that in the investigator’s judgement precludes participation in the study.

Subjects who have received an investigational drug or have used an investigational device in the 30 days preceding study entry.

II.2.J. Tools for assessing endpoints

Clinic assessments will be completed at four different time points during the 8 week study: Days 7, 14, 28 and 56 (± 1 day) and at a fifth time point if tapering-off is required. Subjects will be advised to contact the investigator or site staff should their pain not be controlled and therefore may require additional in-clinic visits. Telephone contacts will be made to subjects on Days 3, 6, and 9 to ensure adequate pain control is achieved through dose titration and that possible side-effects are managed appropriately.

For eligible subjects, the following items will be recorded: standard demographic data; full medical, surgical, and pain medication history; status of OA (including x-ray diagnostic of the diseased joint, ACR Functional Class, OA classification per ACR criteria); the nature, dosage and evaluation of the analgesic treatment of the past month. An x-ray diagnostic of OA taken within the last two years will be acceptable.

Physical examinations will be recorded at the beginning and end of the study. Vital signs will be taken at each visit. Height will be recorded at Visit 1. Weight will be recorded at Visit 1, 3, and 4. All adverse events will be recorded from the first study-related procedure to the last study-related procedure. A statement that the subject meets all eligibility criteria will be documented in the source notes by the Investigator.

The primary objective of this study is to determine the proportion of subjects who experience “moderate”, “good”, or “excellent” pain control during 56 days of treatment with DRUG X. Secondary analyses will include comparisons of the scores from the WOMAC Osteoarthritis Index questionnaire and SF-36 Acute Health Survey (see Appendix B) during 56 days of treatment with DRUG X compared to baseline. Physician and Subject Global Impression of Change scales (7-point scale, see Efficacy Evaluations section) will be done at the Final Visit.

Efficacy Evaluations

Efficacy of DRUG X to treat the signs and symptoms of moderate to severe pain due to OA of the hip or knee will be measured by:

Pain Control Assessment. Subjects will indicate the level of pain control at baseline and during the 14 day treatment period with DRUG X at each visit and at each telephone contact. The question should be asked at approximately the same time of day to ensure consistency. This consists of a five-point evaluation scale from excellent to very poor. For this assessment the subject will be asked: “Think about the pain in your _______ (study joint). Would you rate your pain control today as being: excellent, good, moderate, poor or very poor?”

Functional Status. Subjects will rate their pain, stiffness and physical function at baseline and each visit by means of the WOMAC Questionnaire (Western Ontario and McMaster University Osteoarthritis Index). A one-week recall period will be applied to all questions.
Quality of life. Subjects will complete a 36-item health survey used to evaluate the subject’s physical, social, mental, and general well-being at baseline and each visit by means of the SF-36 Acute Health Survey (see Appendix B). A one-week recall period will be applied to all questions.

Subject/Physician Global Impression of Change. At the completion of the study (or at the early withdrawal visit) the subject and the investigator will answer the question: “Since the start of the study, my [the subject’s] overall target joint status is?” - Very much improved, much improved, minimally improved, no change, minimally worse, much worse, very much worse.

Efficacy Criteria
The primary efficacy parameter of this study is the pain control of the target osteoarthritis hip or knee joint defined as a score of “excellent”, “good”, or “moderate” on the five-point scale: excellent, good, moderate, poor, and very poor. The proportion of subjects with pain control will be given per time point and at endpoint together with a 95% confidence interval. Outcomes are results from the WOMAC Osteoarthritis Index questionnaire, the SF-36 Health Survey and the Physician and Subject Global Impression of Change scores.

Safety Evaluations
All subjects will be considered for the safety evaluation. The incidence of all adverse events will be determined. Special attention will be given to those subjects who have discontinued the trial because of an adverse event, who experienced a severe or serious adverse event or who discontinued the trial due to lack of efficacy. Vital signs and the findings from physical examinations will be assessed.

The following will assess safety:
   a. Vital signs including sitting pulse and blood pressure (after a 5-minute rest), and respiratory rate will be measured at each visit.
   b. Weight will be recorded at Visit 1, 3, and 5 and height will be recorded at Visit 1 only.
   c. A complete medical history will be done at screening only and physical examinations will be done at screening and at Visit 4.
   d. Adverse events will be recorded from the time of the first study-related procedure to the time of the last study-related procedure.

II.2.L. Data analysis method
Efficacy Evaluations
Efficacy analyses will be carried out using the evaluable population. The evaluable population will consist of all subjects who have pain control information at Visit 2 (Week 2). In this case, missing values will be imputed using last observation carried forward (LOCF). Secondary analyses will also be carried out using the observed cases without imputation of missing values. Statistical tests will be carried out at the two-tailed 5% significance level unless specified otherwise.

The primary efficacy parameter will be the proportion of subjects on DRUG X achieving "excellent", "good" or "moderate" pain control compared to placebo on the 5 point scale: excellent, good, moderate, poor and very poor at Week 8 (Day 56). The results will be tabulated and plotted over time. Point estimates and 95% confidence intervals will be provided. A secondary analysis will be carried out using the observed cases.

Secondary analysis will be carried out using the evaluable population as well as the observed cases. Secondary responses include the WOMAC questionnaire, Acute SF-36 Health Survey Quality of Life questionnaire and Physician and Subject Global Impression of Change Scale. For the analyses using the evaluable population LOCF will be used to impute missing instrument scores. Tabulations will include
summary statistics such as the number of observed cases, the mean, standard deviation, minimum and maximum values.

WOMAC scores will be tabulated and plotted over time. Separate results will be tabulated for pain, stiffness, physical functioning and total scores. Significant differences between baseline and Week 8 will be assessed using the paired t-test. Acute SF-36 QoL scores will be tabulated and plotted over time. Significant differences between Baseline and Week 8 will be assessed using the paired t-test. Separate results will be tabulated for total score as well as for sub-scores for physical functioning, physical role limitation, emotional role limitation, social functioning, body pain, general mental health, vitality perception, and general health perception. Global Impression of Change scales provided by investigators and subjects will be tabulated.

**Exploratory Analysis**

The following tabulations and analysis will be presented as part of additional exploratory analyses.

a. A tabulation of the average and final titration doses.

b. A tabulation of the number of acetaminophen 500 mg tablets consumed per week.

c. A tabulation of the Treatment Assessment Questionnaire scores provided by subjects.

d. A comparison of the primary response between subjects taking NSAIDs and a weak opioid/acetaminophen combination and subjects taking only a weak opioid/acetaminophen combination prior to the study will be carried out using the exact Fisher test.

**III. SUGGESTED READINGS**


APPENDIX A. The American College of Rheumatology criteria for the classification and reporting of osteoarthritis

### Classification Criteria for Osteoarthritis of the Hip

**Traditional format**

Hip pain plus at least two of the following:
- ESR of less than 20 mm per hour
- Femoral or acetabular osteophytes on radiographs
- Joint space narrowing on radiographs

**Classification-tree format**

Hip pain plus femoral or acetabular osteophytes on radiographs or hip pain plus joint space narrowing on radiographs and an ESR of less than 20 mm per hour

ESR = erythrocyte sedimentation rate


### Classification Criteria for Idiopathic Osteoarthritis of the Knee

**Traditional format**

Knee pain plus osteophytes on radiographs and at least one of the following:
- Subject age older than 50 years
- Morning stiffness lasting 30 minutes or less
- Crepitus on motion

**Classification-tree format**

Knee pain and osteophytes on radiographs or knee pain plus subject age of 40 years or older, morning stiffness lasting 30 minutes or less and crepitus on motion.

ESR = erythrocyte sedimentation rate

APPENDIX B. Quality of Life (SF-36) Questionnaire

SF-36 ACUTE VERSION

INSTRUCTIONS: This survey asks for your views about your health. This information will help keep track of how you feel and how well you are able to do your usual activities.

Answer every question by marking the answer as indicated. If you are unsure about how to answer a question, please give the best answer you can.

1. In general, would you say your health is:

   (circle one)

   Excellent ................................................................. 1
   Very good .............................................................. 2
   Good ........................................................................... 3
   Fair ............................................................................ 4
   Poor ........................................................................... 5

2. Compared with one week ago, how would you rate your health in general now?

   (circle one)

   Much better now than one week ago ................................................. 1
   Somewhat better now than one week ago ........................................... 2
   About the same as one week ago ....................................................... 3
   Somewhat worse now than one week ago .......................................... 4
   Much worse now than one week ago ................................................ 5

3. The following questions are about activities you might do during a typical day. Does your health now limit you in these activities? If so, how much?

   (circle one number on each line)

<table>
<thead>
<tr>
<th>ACTIVITIES</th>
<th>Yes, Limited A Lot</th>
<th>Yes, Limited A Little</th>
<th>No, Not Limited At All</th>
</tr>
</thead>
<tbody>
<tr>
<td>a. Vigorous activities, such as running, lifting heavy objects, participating in strenuous sports</td>
<td>1</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>b. Moderate activities, such as moving a table, pushing a vacuum cleaner, bowling, or playing golf</td>
<td>1</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>c. Lifting or carrying groceries</td>
<td>1</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>d. Climbing several flights of stairs</td>
<td>1</td>
<td>2</td>
<td>3</td>
</tr>
</tbody>
</table>
e. Climbing one flight of stairs | 1 | 2 | 3
f. Bending, kneeling, or stooping | 1 | 2 | 3
g. Walking more than a mile | 1 | 2 | 3
h. Walking half a mile | 1 | 2 | 3
i. Walking one hundred yards | 1 | 2 | 3
j. Bathing or dressing yourself | 1 | 2 | 3

4. During the past week, have you had any of the following problems with your work or other regular daily activities as a result of your physical health?

(Circle one number on each line)

<table>
<thead>
<tr>
<th>YES</th>
<th>NO</th>
</tr>
</thead>
</table>
a. Cut down on the amount of time you spent on work or other activities | 1 | 2 |
b. Accomplished less than you would like | 1 | 2 |
c. Were limited in the kind of work or other activities | 1 | 2 |
d. Had difficulty performing the work or other activities (for example, it took extra effort) | 1 | 2 |

5. During the past week, have you had any of the following problems with your work or other regular daily activities as a result of any emotional problems (such as feeling depressed or anxious)?

(circle one number on each line)

<table>
<thead>
<tr>
<th>YES</th>
<th>NO</th>
</tr>
</thead>
</table>
a. Cut down on the amount of time you spent on work or other activities | 1 | 2 |
b. Accomplished less than you would like | 1 | 2 |
c. Didn't do work or other activities as carefully as usual | 1 | 2 |

6. During the past week, to what extent has your physical health or emotional problems interfered with your normal social activities with family, friends, neighbours, or groups?

(circle one)

Not at all .................................................................................................................. 1
Slightly ..................................................................................................................... 2
Moderately ............................................................................................................... 3
Quite a bit ............................................................................................................... 4
Extremely ............................................................................................................... 5

304
7. How much bodily pain have you had during the past week?

(circle one)

None ................................................................. 1
Very mild ........................................................... 2
Mild ................................................................. 3
Moderate ........................................................... 4
Severe ............................................................. 5
Very severe ...................................................... 6

8. During the past week, how much did pain interfere with your normal work (including both work outside the home and housework)?

(circle one)

Not at all .......................................................... 1
A little bit ......................................................... 2
Moderately ...................................................... 3
Quite a bit ......................................................... 4
Extremely ...................................................... 5

9. These questions are about how you feel and how things have been with you during the past week. For each question, please give the one answer that comes closest to the way you have been feeling. How much of the time during the past week:

(circle one number on each line)

<table>
<thead>
<tr>
<th>Question</th>
<th>All of the Time</th>
<th>Most of the Time</th>
<th>A Good Bit of the Time</th>
<th>Some of the Time</th>
<th>A Little of the Time</th>
<th>None of the Time</th>
</tr>
</thead>
<tbody>
<tr>
<td>a. Did you feel full of life?</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>6</td>
</tr>
<tr>
<td>b. Have you been a very nervous person?</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>6</td>
</tr>
<tr>
<td>c. Have you felt so down in the dumps that nothing could cheer you up?</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>6</td>
</tr>
<tr>
<td>d. Have you felt calm and peaceful?</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>6</td>
</tr>
<tr>
<td>e. Did you have a lot of energy?</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>6</td>
</tr>
<tr>
<td>f. Have you felt downhearted and low?</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>6</td>
</tr>
<tr>
<td>g. Did you feel worn out?</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>6</td>
</tr>
<tr>
<td>h. Have you been a happy person?</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>6</td>
</tr>
<tr>
<td>i. Did you feel tired?</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>6</td>
</tr>
</tbody>
</table>
10. During the past week, how much of the time have your physical health or emotional problems interfered with your social activities (like visiting friends, relatives, etc.)?

(Circle one)

All of the time.............................................................................................................1
Most of the time .............................................................................................................2
Some of the time ..........................................................................................................3
A little of the time .......................................................................................................4
None of the time...........................................................................................................5

11. How TRUE or FALSE is each of the following statements for you?

(circle one number on each line)

<table>
<thead>
<tr>
<th></th>
<th>Definitely True</th>
<th>Mostly True</th>
<th>Don't Know</th>
<th>Mostly False</th>
<th>Definitely False</th>
</tr>
</thead>
<tbody>
<tr>
<td>a. I seem to get ill a little easier than other people</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
</tr>
<tr>
<td>b. I am as healthy as anybody I know</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
</tr>
<tr>
<td>c. I expect my health to get worse</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
</tr>
<tr>
<td>d. My health is excellent</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
</tr>
</tbody>
</table>
**APPENDIX C. Patient Diary**

1. Please record the number of tablets of EXTRA STRENGTH TYLENOL* you take each day for the pain in your _______________ (study joint). It is best if you can complete this information at the end of each day so you don’t forget to record any tablets you have taken.

Please complete an entry for each day even if you did not require any EXTRA STRENGTH TYLENOL* and write down “0” for those days where you did not take any.

<table>
<thead>
<tr>
<th>Date</th>
<th>Number of tablets of Extra Strength Tylenol taken</th>
<th>Date</th>
<th>Number of tablets of Extra Strength Tylenol taken</th>
</tr>
</thead>
<tbody>
<tr>
<td>DD-MON-YYYY</td>
<td></td>
<td>DD-MMM-YYYY</td>
<td></td>
</tr>
<tr>
<td>Extra days, if applicable</td>
<td></td>
<td>Extra days, if applicable</td>
<td></td>
</tr>
<tr>
<td>Extra days, if applicable</td>
<td></td>
<td>Extra days, if applicable</td>
<td></td>
</tr>
</tbody>
</table>

*Tylenol is a register trademark of McNeil-PPC, Inc.
Chapter 26. Osteoarthritis/arthrosis Long Term Studies: Delay in Structural Progression

Jean-Pierre Raynauld, M.D. ¹
Johanne Martel-Pelletier, Ph.D. ¹
Boulos Haraoui, M.D. ¹
Jean-Pierre Pelletier, M.D. ¹

¹ Osteoarthritis Research Unit
Centre hospitalier de l'Université de Montréal (CHUM)
Hôpital Notre-Dame
Département de Medicine
Université de Montréal
Montréal, Québec
CANADA
I. INTRODUCTORY REMARKS

Osteoarthritis (OA) is disorder which can potentially affect all synovial joints. It is characterized by degeneration of articular cartilage and bone remodelling. The pathological changes can be focal or more generalized and these changes correlate poorly with clinical symptoms and signs. However, it has been suggested that asymptomatic OA, diagnosed radiologically, is a precursor of symptomatic disease. Osteoarthritis, particularly of large joints of the lower limbs is now widely recognized as a major cause of chronic disability in the population. Currently, there are inconsistencies in the classification of drugs for the treatment of OA and the indications for their use.

II. RATIONAL FOR STUDY DESIGN

II.1. Outline of a Typical Development Plan

Multi-center, randomized, double-blind, placebo-controlled, parallel group study in patients with mild to moderate primary knee osteoarthritis according to the American College of Rheumatology (ACR) criteria (Altman 1986, Appendix A), who have been completely withdrawn from their previous analgesics or anti-inflammatory medications or have been newly diagnosed with mild to moderate primary knee OA, and who are not currently taking any analgesics or anti-inflammatory medications. Alternatively, the hip joint may also be used as a target joint to evaluate OA progression. Hip OA is also defined according to specific ACR criteria (Altman 1991, Appendix A).

II.2. Long-Term Studies: Delay in Structural Progression

II.2.A. Objectives

The purpose of the study is to evaluate the efficacy and safety of continuous treatment of subjects with OA of the knee (or hip) over a two (2) year period with DRUG X versus reference DRUG Y (placebo, analgesic, NSAID) in reducing articular cartilage volume loss (or alternatively the joint space width loss) measured as a percentage change from Baseline.

The primary objective is to determine the efficacy and safety of the investigational DRUG X compared to the reference DRUG Y in patients with symptomatic knee (or hip) OA according to the ACR classification having pain more than one month.

II.2.B. Primary Endpoints

Knee is used as a signal joint to assess OA progression

The primary outcome measure is the percentage of cartilage volume loss in the medial compartment (femoral condyle and tibial plateau) from Baseline as assessed by MRI imaging. The method is described in detail in a previous publication (Raynauld 2003). Let the average % cartilage loss in DRUG X and comparator Y treated groups be denoted by X% and Y%, respectively. The parameter of interest is the mean difference of percentage loss between treatment groups relative to the mean percentage loss of the comparator treated group. This parameter is estimated by the ratio, r defined by:

\[ r = \frac{Y\% - X\%}{Y\%} \]

An independent reader will analyse each MRI image and will be blinded to the visit sequence of each image except for the Baseline image.
Alternative primary endpoint for the knee as the signal joint

An alternative endpoint would be to measure the knee minimal joint space width (minimal JSW) in mm on serial standardised radiographs. The weight-bearing posteroanterior film of both knees flexed at 30 degrees (Shuss view) proposed by Piperno et al (Piperno 1998) demonstrate high accuracy and reliability. For this specific view, the patellae of both knees must touch the film cassette, the toes pointed straight ahead vertically relative to the knee and the pelvis touch the table. The angle of knee flexion measured by a goniometer should approximate 30 degrees. With the aid of fluoroscopy, the x-ray beam is adjusted to obtain a horizontal tibial plateau. The interbone distance at the narrowest point (minimal JSW), the joint surface area (JSA), and the mean JSW can be automatically calculated using an image analysis computer and program developed by Cronozier (Cronozier 1995). For the minimal JSW, a loss of 0.17 mm at one year compared to Baseline may be expected.

HIP is used as a signal joint to assess OA progression

For the hip, a pelvic radiograph is obtained annually from patients with a weight-bearing position and standing at 1 meter from the x-ray source with a 20 degree internal foot rotation. Again, the interbone distance at the narrowest point (minimal JSW), the joint surface area (JSA), and the mean JSW can be automatically calculated using the same image analysis computer and program developed by Cronozier (Cronozier 1995).

II.2.C. Secondary Endpoints

Knee is used as the signal joint

Data obtained from the MR acquisitions at 6 months and 1 year will also be analysed in addition to the Baseline and 2 year data to derive a linear rate of the internal compartment cartilage volume change (degradation) over time. This linear change will be contrasted between the two treatment groups.

Besides the internal compartment, MRI volume measurements of the total cartilage, lateral compartment, medial femoral condyle and lateral femoral condyle will also be analysed.

Other secondary efficacy parameters for both knee or hip OA studies:

Change in the Western Ontario McMaster (WOMAC) OA Index (Bellamy 1996) from Baseline to two years of either or all of the 3 following variables:

- a) Pain subscale of the WOMAC
- b) Functional Index of the WOMAC
- c) Total WOMAC Index

Quality of Life Questionnaire (SF-36 Acute) (see Chapter 25, Appendix B)
Use of rescue acetaminophen with a Patient Diary (see Chapter 25, Appendix C)
Patient and Investigator Global Assessment of Efficacy (Visual Analog Scale, 0 to 100 mm, 100 = worse). Number of subjects with an indication for knee or hip replacement surgery, as determined throughout the study.

II.2.D. Study Design

A multi-centre, randomised, double-blind, two-parallel group study comparing the comparator (placebo, analgesic, or NSAID) to DRUG X in subjects with OA of the knee or hip. Following a one week washout period, subjects will be treated with either DRUG X or comparator Y for 24 months.

At Visit 1 (Screening Visit) written informed consent will be obtained prior to any study-related evaluations or procedures. A physical examination, standing x-ray and laboratory tests will be performed and the inclusion and exclusion criteria will be reviewed. Females of childbearing potential (i.e. not post-menopausal or surgically sterilised) will have a urine pregnancy test. The result must be negative for the subject to continue in the study and the subject must agree to use an acceptable method of birth control for
the duration of the study. Eligible subjects will begin a one (1) week wash-out period during which all medications for OA (i.e. NSAIDS and analgesics, with the exception of acetaminophen which will be permitted as rescue medication) will be terminated.

At Visit 2 (Day 0) the inclusion and exclusion criteria will again be reviewed and eligible subjects will be randomised to either DRUG X twice daily or comparator Y twice daily.

At Visit 3 (Month 1) and Visit 4 (Month 2), and at all subsequent visits, subjects will undergo measurement of vital signs, and be assessed for adverse events and changes in concomitant medication (including acetaminophen use). Laboratory tests will be performed at Visit 3 (Month 1), Visit 6 (Month 6), Visit 8 (Month 12), Visit 10 (Month 18) and Visit 12 (Month 24).

At Visit 5 (Month 3), and every three (3) months thereafter, clinical assessments and questionnaires (i.e. SF-36, WOMAC OA Index, Subjects and Investigators Global Assessment of Efficacy and Tolerability) will be completed in addition to measurement of vital signs and assessment for adverse events and concomitant medication use.

MRI and/or radiological assessments will be performed at Visit 2 (Baseline), Visit 8 (Month 12) and Visit 12 (Month 24) of the treatment period. An additional MRI scan will be performed at Visit 6 (Month 6). Measurements of structural parameters will all be carried out centrally by the same evaluators.

Subjects must not use oral or parenteral anticoagulants (with the exception of ASA at a maximum daily dose of 325 mg), oral or topical NSAIDs (other than study medication), immunosuppressive drugs, lithium carbonate, phenytoin, analgesic drugs including over the counter preparations (other than rescue medication provided by the Investigator), other anti-arthritic drugs, including indomethacin, or compounds containing non-approved agents for arthritis or agents claiming to possess disease/structure-modifying properties (e.g. glucosamine and/or chondroitin sulfate containing compounds).

Acetaminophen, provided by the Sponsor, will be the sole analgesic medication permitted during the study. Acetaminophen will be allowed up to a maximum of 4 g daily, and will have to be stopped 24 hours before each study visit. The consumption of acetaminophen, a secondary parameter of efficacy, will be documented in the CRF. Subjects who require more than 4 g daily of acetaminophen must be discontinued from the study.

The medications and other treatments in use for intercurrent illnesses at the time of the Baseline Visit should remain constant for the duration of the study, as evaluated by the Investigator. A subject who is on an established physiotherapy regimen should continue with the same regimen during the study period.

II.2.E. Planned Sample Size for Knee MRI Changes Used as Primary Endpoint
We require 80% power to detect the clinically important difference of r (see primary outcome) as defined in the primary endpoint greater than or equal to 0.3 (30%). The required sample size at the two-sided 0.05 significance level is 110 per group. Assuming an approximate 25% dropout rate, the total sample size is 276 subjects (138 subjects per group).

This sample size estimation is based on the assumption that the comparator Y group will experience a 7.6% reduction of cartilage volume loss of the medial compartment after two years with a standard deviation of 6.0. The DRUG X treated subjects are expected to have a 5.32% reduction of cartilage volume loss after two years with a standard deviation of 6.0.
II.2.F. Study Population
Approximately 280 patients with moderate to severe knee OA will be randomised in this study. A total of 220 patients will be expected to complete the study.

II.2.G. Specific Inclusion Criteria
a. Ambulatory outpatients of either sex between 40 and 80 years of age inclusive and with primary OA of the knee who will not require surgical treatment for at least two years after inclusion.
b. Subjects with OA of the knee meeting the American College of Rheumatology (ACR) classification (see Chapter 25, Appendix A).
c. Subjects complaining of intermittent or constant pain for at least 50% of the time within two months prior to the Baseline Visit (Visit 2) and for whom treatment with NSAIDs is indicated.
d. Subjects with a WOMAC pain subscale index of at least 40 after a 24-hour washout of any analgesics and a seven (7) day washout of any NSAIDs.
e. Subjects with Kellgren and Lawrence Grade 2 or 3 OA of the knee (Kellgren 1957) assessed using a weight-bearing view radiograph taken not more than six (6) months prior to the Baseline Visit (Visit 2). Osteophytes may be present at either the medial or lateral, tibial or femoral margins.
f. The subject must have at least 2 mm and not more than 4 mm of medial joint space width, measured at the narrowest point in the medial compartment, as assessed locally by the Investigator with a caliper.
g. Subjects with at least one of the following three risk factors for increased risk of radiographic progression:
   • Body Mass Index (BMI) > 30
   • Heberden’s Nodes
   • Female gender
h. Subjects capable and willing to give written informed consent prior to enrolment, and who are able to understand and complete the study questionnaires.
i. Female subjects of childbearing potential (i.e. not post-menopausal or surgically sterilised) must have a negative urine pregnancy test at screening and must agree to use an acceptable method of birth control for the duration of the study.

II.2.H. Specific Exclusion Criteria
a. Subjects who have undergone total knee replacement in the contralateral knee within 6 months prior to the Screening Visit (Visit 1).
b. Subjects who have received an intraarticular corticosteroid injection in a lower joint during the three (3) months prior to the Baseline Visit (Visit 2) or any other injection (e.g. hyaluronic acid) within 90 days of the Baseline Visit (Visit 2).
c. Subjects with isolated lateral compartment disease defined by joint space loss in the lateral compartment only.
d. Subjects with OA secondary to a known disorder: Rheumatoid arthritis, seronegative spondylarthropathy, mixed connective tissue disease, collagen vascular disease, psoriasis, inflammatory bowel disease, recently clinically active (within three (3) months) CPPD or crystal-induced arthropathy (e.g. gout), any history of fracture involving the study joint, or any other type of arthritis.
e. Subjects with Class IV functional capacity using the American Rheumatism Association criteria.
f. Subjects who have had surgery in any lower limb joint within 365 days of the Baseline Visit (Visit 2) or arthroscopy, aspiration or lavage in any lower limb joint within 180 days of the Baseline Visit (Visit 2).
g. Subjects who have used indomethacin, or compounds containing non-approved agents for arthritis or agents claiming to possess disease/structure-modifying properties (other than glucosamine and/or chondroitin sulfate containing compounds), in the 14 days prior to the Baseline Visit (Visit 2).
h. Subjects who have used glucosamine and/or chondroitin sulfate containing compounds within:
   • seven (7) days prior to the Baseline Visit (Visit 2) if they have been taking the substance for one (1) month or less;
   • fourteen (14) days prior to the Baseline Visit (Visit 2) if they have been taking the substance for more than one (1) month but less than three (3) months;
   • ninety (90) days prior to the Baseline Visit (Visit 2) if they have been taking the substance for more than three (3) months
i. Subjects who have used medications with MMP-inhibitory properties (e.g. tetracycline or structurally related compounds) within 28 days prior to the Baseline Visit (Visit 2), or took corticosteroids (systemic, >10 days duration) within 28 days of Visit 2.
j. Subjects who require acetaminophen at daily doses >4000 mg (4 g).
k. Subjects who are taking lithium carbonate, phenytoin or anticoagulants (with the exception of ASA up to a maximum daily dose of 325 mg).
l. Subjects who have received chondrocyte transplants in any lower extremity joint.
m. Subjects with comorbid conditions that restrict knee function.
n. Subjects with any significant diseases or conditions, including emotional or psychiatric disorders and substance abuse that, in the opinion of the Investigator, are likely to alter the course of OA or the subject’s ability to complete the study.
o. Subjects with any active acute or chronic infection requiring antimicrobial therapy, or serious viral (e.g. hepatitis, herpes zoster, HIV positivity) or fungal infections.
p. Subjects with a history of a gastrointestinal disorder that could prevent the subject from receiving NSAIDs for the total (2 years) duration of the study. Subjects may, at the Investigator’s discretion, take proton pump inhibitors as required.
q. Subjects with pre-existing malignancy, other than basal cell carcinoma, within ten (10) years of the Screening Visit (Visit 1).
r. Subjects with chronic liver or kidney disease, as defined by AST or ALT 2 times the upper limit of normal (ULN), or by serum creatinine 2.0 mg/dL, at the Screening Visit (Visit 1) and at one repeat testing.
s. Subjects with a known hypersensitivity to DRUG X, NSAIDS (including NSAID comparator) or acetaminophen, or known acetaminophen- or NSAID-induced asthma.
t. Subjects receiving any investigational drug within 30 days prior to the Baseline Visit (Visit 2).
u. Subjects with any contraindications for undergoing MRI.

II.2.1. Tools for Assessing Endpoints

Primary Structural Endpoints
   a. MRI
   b. Standardised Radiograph (Schuss view)

Tools for assessing Secondary Endpoints
   a. WOMAC and SF-36 Questionnaires,
   b. Patient and Physician Global Disease Assessment,
   c. Patient Diary.

Other Endpoints
Blood and urine samples will be collected at the Screening Visit, Visit 6 (Month 6), Visit 8 (Month 12), Visit 10 (Month 18) and Visit 12 (Month 24). The following laboratory tests will be performed:
   a. Haematology: haemoglobin, haematocrit, white blood cell count (WBC) (total and differential), red blood cell count (RBC), platelet count and erythrocyte sedimentation rate (ESR).
   b. Chemistry: creatinine, urea, aspartate aminotransferase (SGOT/AST), alanine aminotransferase (SGPT/ALT), gamma-glutamyltransferase (gamma-GT), alkaline phosphatase, total bilirubin, sodium, potassium and uric acid.
c. Urinalysis: pH, protein, glucose, ketone bodies, leukocytes, nitrite, haemoglobin/erythrocytes, urobilinogen and bilirubin. If haemoglobin/erythrocytes or protein are found in the urine, a microscopic investigation will be performed.

d. Urine pregnancy test will be performed on female subjects of child bearing potential at Visit 1 only. The urine pregnancy test must be negative in order for females of child bearing potential to proceed in the study. Females are considered of child bearing potential unless they are post-menopausal or surgically sterilised.

Frozen sera and urine will be collected at Visit 2 (Month 1), Visit 6 (Month 6), Visit 8 (Month 12), Visit 10 (Month 18) and Visit 12 (Month 24). Frozen samples will be stored until the end of the study and batch-shipped to the laboratory designated by the Sponsor. Ideally, samples will be stored at -70 degrees C.

All laboratory tests will be performed by the central laboratory. A copy of the laboratory report will be collected by the Sponsor.

**Safety Considerations**

All adverse events (AEs), whether observed by the Investigator or reported by the subject, will be recorded in the CRF provided by the Sponsor. The date of onset, duration, intensity, action taken due to the event, outcome, and relationship to the study medication will be indicated.

The severity of the event will be classified according to the following terms:

a. mild: symptom is manifest but is tolerated
b. moderate: normal activity affected
   c. severe: severe effect or inability to work or necessary to discontinue the study medication

The causality will be classified as:

a. probable
b. possible
c. no causal relationship
d. unclassified

For all AEs, the Investigator must obtain all information sufficient to determine whether the event meets the criteria for SAE reporting, and to determine the causality and outcome of the AE. Follow-up of all AEs reported during the study is required until the event resolves or stabilises at a level considered as acceptable by the Investigator and the Sponsor.

**II.2.J. Specific Criteria for early Withdrawal and Discontinuation**

Subjects may withdraw from the study at any time and for any reason. The Investigator may choose to withdraw a subject from the study if, in the opinion of the Investigator, continued participation in the study may compromise the safety or well-being of the subject. In the event of premature discontinuation from the study, the Investigator should determine the primary reason for discontinuation. A final evaluation should be completed at the time of discontinuation.

In the event that a subject discontinues the study at any time after randomisation, the Investigator will determine the reason for the discontinuation. Subjects who discontinue study medication will proceed to the Visit 12/End of Treatment visit and all assessments required for that visit will be performed.

If the subject discontinues from the study prior to Month 8 (Visit 12) but is willing to continue being followed in the study up to Month 12, he/she will undergo all assessments required at Visit 12/End of Treatment except the MRI and Schuss view knee radiographs unless these would have been done at the
subject’s next scheduled study visit. Subjects will be treated for OA at the Investigator’s discretion and will undergo the following procedures at Visit 8 (Month 12) only:

a. MRIs
b. Schuss view knee radiographs.
c. Vital signs (including weight, blood pressure and pulse measured after five (5) minutes at rest in the seated position).
d. If the subject is unwilling to continue in this manner or, in the opinion of the Investigator, it is contrary to the well-being of the subject to continue, then a final study visit will be done at which all assessments required at the Visit 12/End of Treatment Visit will be completed.

Subjects who withdraw more than one month after randomisation will not be replaced.

II.2.J. Data Analysis Method

Intent-To-Treat Subjects
A subject who is randomised to treatment and takes at least one dose of study medication will be included in the Intent-To-Treat population.

Per Protocol Subjects
A subject who completes the trial will be included in the Per Protocol population.

Analysis of Efficacy
The primary objective of this study is to evaluate the efficacy and safety of continuous treatment over a two (2) year period with DRUG X versus comparator Y in reducing articular cartilage volume loss measured as a percentage change from Baseline.

The primary efficacy assessment is based on the ITT sample size of all 276 subjects. Assume 80% of subjects in each group complete the trial (with expected % medial compartment cartilage losses equal to 7.6% and 5.32% in the comparator Y and DRUG X groups, respectively, at two years). Also assume 20% of subjects per group average only 1 year of study and are lost to follow-up (with % cartilage losses averaging 3.91% and 2.74% in the comparator Y and DRUG X groups, respectively). Then 134 subjects per group would be required to be followed, to achieve the overall 0.05 significance level with 80% power. Hence, the total sample size of 276 subjects is sufficient to meet the primary study objectives using the ITT group.

Baseline Values and Demographics
Characteristics of the subjects such as age, gender and race in each treatment group will be compared (including comparisons by centre), using the appropriate parametric tests. Efficacy variables assessed at Visit 2 are considered as the Baseline values and will also be compared.

Stratification is not used for the primary efficacy analysis. The design uses a completely randomized assignment plan. However, secondary analyses will be performed adjusting for influential covariates (using e.g.: the Analyses of Covariance) should they be out of balance at Baseline. The ANACOVA may also be used, secondarily, to investigate site differences.

Analytical Plan
Analysis of Primary Efficacy Outcome:
The statistic, \( S = \ln (1-r) \), where \( r \) is defined in the primary outcome section, will be used to test the primary hypothesis, which can be stated as follows:

Hypothesis: The expected value of S will be less than or equal to \( \ln (.7) \)
Statistical Method: The comparison of treatment groups will use the normal approximation for the distribution of S. The comparison is based on the standardised value of S.

Descriptive Efficacy Analyses:
Standard analyses will include tabulations of means, percentages, standard deviations and confidence intervals. Data listings, and plots, will also be presented when useful. Analyses by visit will compare the two treatment groups on both a Per Protocol and an Intent-To-Treat basis.

The average of continuous variables, such as the percentage of cartilage volume loss, rescue use of acetaminophen, WOMAC OA index and all VAS scores will be analysed using standard normal theory.

Nominal variables, such as the number of subjects with an indication for knee replacement surgery will be analysed using Yates’ continuity corrected chi-square test.

Missing Value Strategies
To assess the sensitivity of the results to the presence of missing data, any missing scores will be imputed using two strategies. The last observation carried forward and at a given time point, the last available observation for the subject plus the average change from the time of last observation for the subject computed using subjects in the same treatment group who have non-missing observations.

Analysis of Safety
Randomised subjects receiving at least one dose of study medication will be included in the safety analyses. Safety analyses will be performed in terms of incidence and severity of adverse and/or unexpected events. These will be tabulated and compared between treatment groups. In addition, a complete listing of all reports of adverse and/or unexpected events will be presented.

Sitting blood pressure and heart rate will be summarised at each visit. Means, standard deviations, minimum and maximum individual values will be presented and compared.

Newly occurred abnormalities in the laboratory values and ECG results will be tabulated for each group. The proportion of subjects reporting such experiences will be compared.

Safety Monitoring Committee
An independent safety monitoring committee, not otherwise involved in the conduct of the trial, will be the primary data and safety advisory group for the Sponsor. The safety monitoring committee will review the results of the interim analysis and will periodically review study results, evaluate the treatments for excess adverse effects, determine whether the basic trial assumptions remain valid, and judge whether the overall integrity and conduct of the trial remain acceptable.

III. EXAMPLES OF LANDMARK WELL DESIGNED TRIALS


IV. SUGGESTED READINGS


Chapter 27. Rheumatoid Arthritis

Jean-Pierre Raynauld, M.D. ¹

Johanne Martel-Pelletier, Ph.D. ¹

Boulos Haraoui, M.D. ¹

Jean-Pierre Pelletier, M.D. ¹

¹ Osteoarthritis Research Unit
Centre hospitalier de l’Université de Montréal (CHUM)
Hôpital Notre-Dame
Département de Medicine
Université de Montréal
Montréal, Québec
CANADA
I. INTRODUCTORY REMARKS

Rheumatoid arthritis (RA) is a systemic inflammatory disease characterised by a proliferating synovitis leading to cartilage and bone destruction, and resulting in joint deformities and increasing functional limitations. Rheumatoid arthritis affects about 1% of the adult population. It occurs 2 to 3 times more frequently in women than in men. A great proportion of RA patients will rapidly develop major disabilities and almost 50% will experience work loss within ten years of diagnosis (1). Progression of the disease may also lead to premature death (2).

II. STUDIES FOR REGISTRATION OF NEW DRUGS

II.1. Outline of a typical plan

Evidence is ample that joint damage is an early phenomenon and progresses relentlessly over the years. Moreover, there is a direct causal link between the synovitis, the anatomical damage and disability (3). A general consensus has emerged that the key goals of therapy in RA are early rapid control of joint inflammation and prevention of joint destruction.

Therefore, a number of outcomes can be evaluated in clinical trials of RA (4-5):

a. Reduction in signs and symptoms
b. Improvement in functional disability
c. Prevention of anatomical radiographic damage
d. Safety issues
e. Considerations for short-term symptom modifying trials

Reduction in the signs and symptoms of RA
The goal is to demonstrate an improvement in signs of disease activity as well as symptoms.

Ordinarily, trials should be of a duration of at least six months, unless the product belongs to an already well-characterised pharmacological class (e.g., NSAIDs) for which trials of three-month periods are sufficient to establish efficacy for signs and symptoms. Six-month trials are desirable because RA is a disease of long duration. Interventions which provide only short-term, time-limited, benefit are unlikely to have overall value to patients.

In evaluating signs and symptoms, methods which evaluate response over time are preferable to methods which incorporate only the baseline value and the final observation, unless there is a reason to weigh last-visit symptoms rather than intermediary symptoms.

Several individual clinical or laboratory outcome measures can be used, but it has become the norm to use validated composite endpoints or indices of signs and symptoms. These composites may also be used to construct categorical endpoints for patient success or failure. For example, the Paulus criteria (6) or the more widely accepted American College of Rheumatology (ACR) definition of improvement (ACR 20, 50, 70) (7,8) are now the norm used to assess a patient’s response. ACR 20, 50 or 70 refers to an improvement of 20%, 50% or 70% in:

a. The number of swollen joints and
b. The number of tender joints and
c. 3 of the following measures:
d. Patient’s evaluation of pain on a visual analogue scale (VAS)
e. Patient’s global assessment on a VAS
f. Physician’s global assessment on a VAS
g. Health Assessment Questionnaire (HAQ) (see appendix 1)
h. Erythrocyte sedimentation rate (ESR) or C-Reactive Protein (CRP)

Example: Success for each patient in a six-month trial could be predefined as meeting the criteria for improvement over baseline in at least four of six monthly observations and not dropping out because of toxicity.

The use of the 66- or 28-joint count is appropriate for the evaluation of the swollen and tender joints (9).

A Major Clinical Response can also be defined as a continuous six-month period of success by the “ACR 70”. Therefore the trial duration should last a minimum of seven months for an agent expected to have a rapid onset of action, and longer for agents with less prompt effects.

A Complete Clinical Response or Remission refers to a Major Clinical Response coupled to radiographic arrest based on validated X-ray scoring systems (10,11). Complete clinical response connotes a benefit requiring ongoing drug therapy while remission is defined by the same result while off all antirheumatic drugs. The 1981 ACR remission criteria (12) require at least five of the following: morning stiffness less than 15 minutes, no fatigue, no joint pain by history, no joint tenderness or pain on motion, no swelling of joints or tendon sheaths, and erythrocyte sedimentation rate (ESR) less than 20 for males or less than 30 for females. The duration of the trial should be of at least one year or longer, depending on the rapidity of onset of the drug. Trials evaluating a complete clinical response should use a categorical endpoint (patient complete response or treatment failure) as the primary outcome measure.

Another validated composite score is the Disease Activity Score (DAS). The original DAS combines the Ritchie articular index, 44 counts of swollen joints, ESR and an assessment of the patient’s global health (13). The DAS28 is a validated subscale combining 28 swollen joint counts, 28 tender joint counts, ESR and the patient’s global health (14). The DAS28 is a continuous measure that was developed to assess the level of disease activity at a certain point in time (status score), which gives it an added value compared to the ACR response scores. However, its complicated formula requires the use of a computer or hand-held calculator, which are readily available. The DAS has also been validated against the ACR response criteria; therefore, these 2 validated measures can be used interchangeably when evaluating response (15,16).

Because most RA outcome measures have a high degree of subjectivity, the highest confidentiality inpatient and assessor blinding should be sought to achieve a credible inference. Therefore all clinical evaluations, including joint counts, should be assessed by an independent party with no knowledge of the subject’s history (blinded assessor).

**Prevention of disability**

Currently, the Health Assessment Questionnaire (HAQ) (17) (see appendix 1) and the Arthritis Impact Measure Scales (AIMS) (18) (http://www.qolid.org/public/aims/cadre/cadre.htm) are adequately validated measures for use as the primary outcome measure in assessing disability. Studies should be of a duration of at least six months, but interventions/drugs seeking this claim should be two to five years in duration and must concomitantly demonstrate improvement in signs and symptoms. Since the full effect of RA on a patient is not captured without the use of more general HR-QOL measures, a validated measure such as the SF-36 should also be collected and patients should not worsen on these measures over the duration of the trial (19, SF-36 Web page http://www.sf-36.com)(Appendix 2).

**Prevention of structural damage**

Prevention of structural damage is an important goal of RA therapy. Trials evaluating this outcome should be at least one year in duration and based on comparisons of films taken at one year (and subsequent yearly points) with those taken at baseline. All randomised patients should have films at both time points,
regardless of whether they are continuing treatment. Patients dropping out of the trial should have films taken at that time. Pre-specification of the handling of dropouts is especially important in these trials.

Different outcome measures can be used, as follows:

a. Slowing X-ray progression using the Larsen, the modified Sharp or another validated radiographic index (10,11);
b. Prevention of new X-ray erosions or maintaining an erosion-free state; and
c. Other measurement tools, such as MRI or ultrasonography, could be employed. However, because of the potential of the technique to identify small albeit statistically significant changes, the magnitude of the difference that would reflect actual patient benefit is unclear and needs to be established.

Because slowing of radiographic progression does not in itself define a direct patient benefit, it is expected that the agent would also demonstrate efficacy in one of the other claims (e.g., prevention of disability). However, some agents are not intended to affect acute inflammation but are designed to prevent or slow joint destruction by other means. Since the ultimate goals of slowing joint destruction are to improve symptoms and/or to preserve functional ability, slowing radiographic progression of disease is considered a surrogate marker for overall patient benefit in RA.

Safety issues
Every RA investigational therapy raises safety concerns. Whenever there is a potential for significant toxicity, long-lasting or delayed-onset, it is desirable to design the Phase 2 studies to provide a group of patients with longer term follow-up preceding the larger Phase 3 studies. Provisions for long-term follow-up can also be helpful in addressing issues of immunosuppression, opportunistic infections, neoplasia, and induction of autoimmune disease. Standard toxicity grading scales and stopping rules are also desirable in Phase 2.

Considerations for short-term symptom modifying trials
Phase 2 trials are used to better define the dose- and exposure-related activity and toxicity of the agent. Enough information should be generated to ensure that the Phase 3 trials are conducted safely and with a probability of success. Once a reasonably safe range of doses has been established, randomised, parallel-arm dose-comparison trials are usually recommended. The use of a placebo arm is desirable for several reasons. First, if no difference is found among doses, there is usually no other way to determine whether all doses were equally effective or equally ineffective. Second, if a dose-response trend is found, the placebo arm may indicate the possible magnitude of the observed effect. If use of a placebo is impossible, designs should include wide dose ranges or durations, or repetitions. Active controlled designs that specify an arm with a well-characterised known therapy can also be very useful.

The overall goal of Phase 3 work is to demonstrate the efficacy of the product in convincing controlled trials and to accrue a sufficient safety database. The new drug is tested against an accepted marketed drug in a randomised (frequently placebo-controlled) double blind trial.

II.2. Short-term trial in RA

II.2.A. Objectives
The primary objective is:

a. to evaluate the clinical efficacy of drug X versus placebo in the treatment of signs and symptoms of RA.

The secondary objectives are:

b. to evaluate the clinical efficacy of drug X versus drug Y based on pain intensity;
c. to determine an effective dose-range for drug X in the treatment of the signs and symptoms of RA, and enable dose selection for subsequent studies;
d. to evaluate the safety and tolerability of drug X in patients with RA;
e. to evaluate health outcomes data by using the EuroQol Questionnaire; and
f. to characterise the population pharmacokinetics of drug X in subjects with RA and assess the presence of a pharmacokinetic/pharmacodynamic relationship with clinical outcome.

II.2.B. Primary endpoints
a. The percentage of ACR20 responders

II.2.C. Secondary endpoints
a. Change from baseline in tender joint count (68 joints) at each scheduled visit;
b. Change from baseline in swollen joint count (66 joints) at each scheduled visit;
c. Change from baseline in physician’s global assessment of arthritis condition at each scheduled visit;
d. Change from baseline in patient’s global assessment of arthritis condition at each scheduled visit;
e. Change from baseline in functional disability index (HAQ) at each scheduled visit;
f. Change from baseline in CRP at each scheduled visit;
g. Percentage of patients discontinuing due to lack of efficacy; and
h. Average total daily dose of rescue medication.

II.2.D. Study design
The study will be a multicentre, randomised double-blind placebo and active controlled parallel group dose ranging and will evaluate the safety and efficacy of 3 oral doses of drug X (A mg, B mg and C mg) relative to placebo and drug Y (D mg) in subject with clinically active rheumatoid arthritis treated for 24 weeks.

All concomitant medications taken during the study will be recorded in the CRF with indication, dose, route of administration as well as the start date and the end date. Medication allowed and prohibited is specified below:

a. Rescue analgesic therapy: up to 3,000 mg of acetaminophen per day is allowed for pain relief. However, rescue medication is not allowed 12 hours prior to a clinic visit in order to minimise the effect of acetaminophen on study endpoints; and
b. Prohibited medications:
   • Gastroprotective agents: misoprostol, sulcrafate
   • Proton pump inhibitors
   • H2 blockers
   • Any analgesic or NSAID other than those allowed by the protocol
   • Oral corticosteroid (equivalent to > 10 mg of prednisone) or initiated within 4 weeks of baseline
   • Intra-articular injections

II.2.E. Planned sample
Sample size assumptions: the primary endpoint is the proportion of ACR 20 responders at 24 weeks. Assuming a drug X response rate of 45% and a placebo response rate of 25%, a sample size of XXX evaluable subjects per treatment group is sufficient to detect a 20% difference between the 2 groups with 90% power and a 5% significant level. Assuming a 20% dropout rate between randomisation and Week 24, a total of YYY subjects is required to be randomised in the study.

II.2.F. Study population
A sufficient number of male and female patients with ACR-defined RA will be screened for enrollment in order to randomise XXX subjects (XXX/5 per treatment group). The study will be conducted at centres in YYYY countries.

II.2.G. Inclusion criteria
a. Males and females of 18 years of age or older;
b. Fulfillment of the 1987 American College of Rheumatology (ACR) criteria for RA (Appendix E) with a disease duration ≥ 6 months;
c. Active disease defined by the presence of the following criteria (based on 66/68 joint counts):
   - 5 or more swollen joints at screening and baseline visit
   - 6 or more tender joints at screening and baseline visit
d. Subjects receiving oral corticosteroids must be receiving a stable dose equivalent to prednisone ≤ 10 mg/day for at least 12 weeks prior to screening;
e. Subjects who are currently receiving DMARD therapy (including methotrexate, sulfasalazine, hydroxychloroquine), must be on stable dose for at least 8 weeks prior to screening;
f. Women of child-bearing potential must use adequate contraceptive precautions (abstinence, oral contraceptives, barrier and spermicide, IUD or surgical sterilisation); and
g. Before any study-specific procedure, subjects must give informed consent for participation in the study.

II.2.H. Exclusion criteria
a. Known history of hypersensitivity or intolerance to NSAIDs, aspirin, COX-2 inhibitors or acetaminophen;
b. ACR functional class IV;
c. Receipt of any investigational drug within 28 days or five half-lives (whichever is longer) of study drug initiation;
d. Any clinical or laboratory abnormality found at screen or baseline which, in the opinion of the investigator, is clinically significant and would compromise the conduct or outcome of the study;
e. History of neoplasm or lymphoproliferative disease including lymphoma. Removed basal cell carcinoma is acceptable;
f. Active infection requiring antibiotic therapy;
g. Intra-articular, soft tissue, or intra muscular corticosteroid injections during the 4 weeks prior to screening;
h. Initiation or change of dose of a standard DMARD within 12 weeks prior to baseline;
i. Use of oral corticosteroids at doses greater than the equivalent of 10 mg/day of prednisone or initiation of treatment within 4 weeks prior to baseline;
j. Initiation of, or change to, an established physiotherapy program within 2 weeks prior to baseline. An established program must continue unchanged throughout the study period;
k. Any disorder which compromises the ability of the subject to give written consent and/or to comply with study procedures; and
l. Any other condition which, in the investigator’s opinion, would preclude the subject from participating.

II.2.I. Tools for assessing endpoints

Analysis of primary efficacy endpoint
The primary efficacy endpoint is the percentage of ACR 20 responders (See Section II.1.a above, “Reduction in the signs and symptoms of RA”). Primary inferences will be based on the ITT population at Week 24, the last observation carried forward (LOCF) endpoint. Initially, an overall chi-square test will be performed to determine any drug X effect versus placebo. If this test is positive at the 5% significance level, the differences between each dose level of drug X will be investigated using logistic regression techniques adjusting for co-variates. In addition, dose response of drug X will be investigated.

Analysis of secondary efficacy endpoints
Each of the secondary variables will be analysed using the final model taken from the primary analysis. All secondary variables will be analysed as drug X versus placebo, and differences between drug X and drug Y will be investigated. Continuous secondary efficacy variables will be analysed using analysis of covariance. Results will be presented as adjusted means of 95% confidence intervals between each dose level of drug X and placebo. Binary efficacy endpoints will be analysed using logistic regression
techniques, adjusting for appropriate covariates. Results will be presented as adjusted odds ratios and 95% confidence intervals around the odds ratios.

Safety
Adverse events (AEs) will be coded using MedDRA. A summary of the number and percentage of subjects with the following adverse events will be displayed by treatment group: all AEs, drug-related AEs, serious AEs, and AEs leading to permanent discontinuation of the study drug.

Each laboratory value will be flagged to show whether it is within, below or above the normal range, and if it is of a clinical concern.

II.2.J. Specific criteria for early withdrawal and discontinuation
Randomised subjects who discontinue from the study prematurely will not be replaced. However, subjects who fail the screening period (screen failures) will be replaced.

If a subject is prematurely discontinued from participation in the study for any reason, the investigator must make every effort to perform the following procedures: completion of ACR20 assessments, vital signs, haematology and biochemistry laboratory assessments, and adverse events assessment.

II.3. Long-term disease modifying trials
Trials aiming to prevent structural radiographic damage should be of a duration of at least one year, with a second year extension to show sustainability of effect.

The study is designed as a one-year superiority trial to show efficacy of drug X in combination with MTX in patients with active RA and inadequate response to MTX, and maintained on MTX.

II.3.A. Objectives
The primary objective is:

a. to compare the clinical efficacy of drug X used in combination with MTX versus MTX alone.

The secondary objectives are:

a. structural damage progression;

b. the proportion of patients with a Major Clinical response, and

c. to evaluate the safety and tolerability of drug X in patients with RA.

II.3.B. Primary endpoints

a. symptomatic relief as measured by the ACR20 response at six months;

b. functional improvement as measured by the HAQ at one year, and

c. structural radiographic damage as assessed by the total modified Sharp score at one year.

II.3.C. Secondary endpoints

a. ACR 50 and 70 response at six months;

b. ACR 20, 50 and 70 response at one year;

c. Radiographic erosion and joint space narrowing score assessed by the modified Sharp method at one year;

d. The proportion of patients with a Major Clinical response (ACR 70 for six continuous months);

e. The disease activity measured by the DAS 28 at six months and one year;

f. The functional improvement at six months measured by the HAQ, and

g. The safety of drug X.
II.3.D. Study design
The study will be a multicenter, randomised double-blind placebo controlled parallel group of one-year duration to evaluate the safety and efficacy of drug X, relative to placebo in subjects with clinically active rheumatoid arthritis despite adequate background therapy with methotrexate. A second year extension is planned to gather further safety data.

Concomitant therapy:
- a. Subjects will continue their current MTX dose unchanged up to Month 6; during this period, only decreases due to toxicity will be allowed;
- b. All other DMARDs (with the exception of MTX) are not permitted during that period;
- c. NSAIDs including ASA are permitted, provided the dose is stable up to Month 6; decreases in NSAIDs dose are permitted only in case of toxicity;
- d. Oral corticosteroids (equivalent to > 10 mg of prednisone) are permitted provided the dose is kept stable during the first 6 months of study;
- e. Rescue analgesic therapy: acetaminophen – a combination of acetaminophen and narcotics may be used except 12 hours prior to a joint evaluation visit, and
- f. Intra-articular injections should be avoided; however, if necessary, up to 2 injections are permitted during the first 5 months. No injections are allowed between Months 5 and 6.

II.3.E. Planned sample
Sample size assumptions: The group receiving drug X in combination with MTX will be compared to the placebo control group receiving MTX alone. Sample size will be based on a 5% level (2 tailed) of significance. A total of XXX randomised in a 1:1 ratio will yield a 99% power to detect a 20% difference in ACR20 response between the two groups. Based on hierarchical testing procedure for the co-primary endpoints, this sample size will allow the detection of 18% difference in the HAQ response and a 60% difference in the total modified Sharp score with a power of 90%. These power calculations are based on the results observed in the Phase 2 trials.

II.3.F. Study population
A sufficient number of male and female patients with ACR-defined RA and fulfilling the Inclusion/Exclusion criteria will be screened for enrollment in order to randomise XXX subjects (XXX/2 per treatment group). The study will be conducted at centres in YYYY countries.

II.3.G. Specific inclusion criteria
- a. Males and females of 18 years of age or older;
- b. Fulfillment of the 1987 American College of Rheumatology (ACR) criteria for RA (Appendix E) with a disease duration ≥ 6 months;
- c. Active disease defined by the presence of the following criteria (based on 66/68 joint counts):
  - 10 or more swollen joints at screening and baseline visit
  - 12 or more tender joints at screening and baseline visit
  - CRP ≥ 1 mg/dL at screening
- d. Subjects receiving oral corticosteroids must be receiving a stable dose equivalent to prednisone ≤ 10 mg/day for at least 12 weeks prior to screening;
- e. Subjects must have been taking MTX for at least 3 months at a dose of ≥ 15 mg per week and be on a stable dose for the last 4 weeks;
- f. All other DMARDs with the exclusion of MTX must have been discontinued 4 weeks prior to baseline. In the case of infliximab, subject must have discontinued treatment at least 8 weeks prior to baseline;
- g. Women of child-bearing potential must use adequate contraceptive precautions (abstinence, oral contraceptives, barrier and spermicide, IUD or surgical sterilisation);
- h. Before any study-specific procedure, subjects must give informed consent for participation in the study.
II.3.H. Specific exclusion criteria

a. Women who are pregnant or breast-feeding;
b. Women of child-bearing potential unwilling or unable to use an acceptable method of birth control;
c. ACR functional class IV;
d. Receipt of any investigational drug within 28 days or 5 half-lives (whichever is longer) of study drug initiation;
e. Any clinical or laboratory abnormality found at screen or baseline which, in the opinion of the investigator, is clinically significant and would compromise the conduct or the outcome of the study;
f. History of neoplasm or lymphoproliferative disease including lymphoma; removed basal cell carcinoma is acceptable;
g. Active bacterial infection requiring antibiotic therapy;
h. Active TB requiring treatment within the previous 3 years; subjects with a positive PPD at screening need to have completed treatment for latent TB and to have a normal chest X-ray;
i. Subjects with a known history of positivity for HIV, Hepatitis C and Hepatitis B; subjects with herpes zoster which resolved less than 2 months prior to enrolment;
j. Intra-articular, soft tissue, or intra-muscular corticosteroid injections during the 4 weeks prior to screening;
k. Use of oral corticosteroids at doses greater than the equivalent of 10 mg/day of prednisone or initiation of treatment within 4 weeks prior to baseline;
l. Any disorder which compromises the ability of the subject to give written consent and/or to comply with study procedures, and
m. Any other condition which, in the investigator’s opinion, would preclude the subject from participating.

II.3.I. Tools for assessing endpoints

Several individual clinical or laboratory outcome measures can be used to assess the changes in signs and symptoms, but it has become the norm to use validated composite endpoints or indices of signs and symptoms. These composites may also be used to construct categorical endpoints for patient success or failure. For example, the Paulus criteria (6) or the more widely accepted American College of Rheumatology (ACR) definition of improvement (ACR 20, 50, 70) (7,8) are now the norm used to assess a patient’s response. ACR 20, 50 or 70 refers to an improvement of 20%, 50% or 70% in:

a. The number of swollen joints and
b. The number of tender joints and
c. 3 of the following measures:
d. Patient’s evaluation of pain on a visual analogue scale (VAS)
e. Patient’s global assessment on a VAS
f. Physician’s global assessment on a VAS
g. Health Assessment Questionnaire (HAQ) (see appendix 1)
h. Erythrocyte sedimentation rate (ESR) or C-Reactive Protein (CRP)

Disability is assessed by using the Health Assessment Questionnaire (HAQ) (17) (see appendix 1) and the Arthritis Impact Measure Scales (AIMS) (18) (http://www.qolid.org/public/aims/cadre/cadre.htm). Since the full effect of RA on a patient is not captured without the use of more general HR-QOL measures, a validated measure such as the SF-36 should also be collected and patients should not worsen on these measures over the duration of the trial (19, SF-36 Web page http://www.sf-36.com)(Appendix 2).

The progression of structural damage is assessed by:

a. Slowing X-ray progression using the Larsen, the modified Sharp or another validated radiographic index (10,11);
b. Prevention of new X-ray erosions or maintaining an erosion-free state; and
c. Other measurement tools, such as MRI or ultrasonography, could be employed. However, because of the potential of the technique to identify small albeit statistically significant changes, the magnitude of the difference that would reflect actual patient benefit is unclear and needs to be established.

II.3.J. Specific criteria for early withdrawal and discontinuation
Randomised subjects who discontinue prematurely from the study will not be replaced. However, subjects who fail the screening period (screen failures) will be replaced.

Study therapy must be immediately discontinued for the following reasons:
  a. Withdrawal of consent;
  b. Any clinical or laboratory adverse event or intercurrent illness which, in the opinion of the investigator, indicates that continued treatment is not in the best interest of the patient;
  c. Inability of the patient to comply with the requirements of the protocol;
  d. Pregnancy; and
  e. Use of prohibited or restricted medications.

If, for any reason, a subject is prematurely discontinued from participation in the study, the investigator must make every effort to perform the following procedures: completion of ACR assessments, DAS evaluation, vital signs, haematology and biochemistry laboratory assessments, adverse events assessment and X-rays.

II.3.K. Data analysis method

*Analysis of primary efficacy endpoints.*
Co-primary analysis includes, in the order of sequential testing, comparisons between drug X and placebo in:
  a. ACR20 response at 6 months by chi-square test with correction for continuity. Percentage improvement in individual core components of the ACR response will be summarised by treatment;
  b. Functional improvement measured by the HAQ; a HAQ response is defined as a reduction of 0.3 unit from baseline; comparisons between treatment groups will be performed with a continuity chi-square test; and
  c. Structural radiographic damage progression measured by the total modified Sharp score; a non-parametric analysis of covariance model (ANCOVA) will be used to compare the changes from baseline of the scores.

Subjects who discontinue the study prematurely will be considered non-responders at all scheduled protocol visits subsequent to the point of discontinuation in the analysis of the ACR and the HAQ responses.

*Analysis of secondary efficacy endpoints:*
  a. The number and percentage of patients achieving a Major Clinical Response in each therapeutic group will be summarised for data collected up to Month 12 and the 95% confidence interval will be computed for the treatment difference; and
  b. Changes from baseline in the DAS28 score at Months 6 and 12 will be summarised by treatment arm, and the 95% confidence interval will be computed for the treatment difference.

*Safety analysis*
Adverse events (AEs) will be coded using MedDRA. A summary of the number and percentage of subjects with the following adverse events will be displayed by treatment group, as follows: all AEs, drug-related AEs, serious AEs, and AEs leading to permanent discontinuation of study drug.
Each laboratory value will be flagged to show whether it is within, below or above the normal range, and if it is of a clinical concern.

### III. SPECIAL CONSIDERATIONS FOR JUVENILE RHEUMATOID ARTHRITIS

Juvenile Rheumatoid Arthritis (JRA) is a heterogeneous group of diseases that share the common feature of chronic, idiopathic synovitis, with onset prior to 16 years of age. These disorders have been divided into clinically distinct subsets based on the extent of joint involvement and extra-articular manifestations: pauci-articular, poly-articular, and systemic-onset JRA as well as oligoarthritis associated with HLA-B27, and they have been further subdivided based on clinical courses (20). Immunogenetic subsets appear to correlate with these clinical course subsets and are also distinct from adult RA (21). Of these various entities, polyarticular JRA is similar in many aspects to adult RA, particularly in clinical signs and symptoms, synovitis, and similar efficacy responses to some existing pharmacotherapy (NSAIDs, methotrexate, and prednisone). The application of principles in the conduct of clinical trials for adult RA largely applies to JRA as well, and this section outlines only those areas of difference from adult RA. The Committee on Drugs of the American Academy of Paediatrics has published guidelines for the ethical conduct of studies to evaluate drugs in paediatric populations (22) as well as general considerations for the clinical evaluation of drugs in infants and children (23).

As a general principle, children should not be subjected to an agent that has not been first tested for safety in adults. Testing may begin in children, but only when the anticipated benefits based on existing knowledge justify the anticipated risks. An agent developed specifically for use in JRA (e.g., a biological agent targeted against a specific pathogenic process that is unique to JRA and not present in adult RA) may need to be tested first in children as exposure in adult RA patients or even normal adult volunteers may be unrevealing.

#### III.1. Reduction in the signs and symptoms of JRA

All JRA trials should evaluate improvement based on a validated endpoint for improvement. Currently, the one validated approach is the definition of improvement established by the JRA core set: three of six (MD global, parent/patient global, number of active joints, number of joints with limited range of motion, functional ability, and ESR) improved by at least 30% and no more than one of six worsened by more than 30% (24). Because the JRA definition of improvement was validated using a trial of methotrexate, which primarily included polyarticular JRA patients, protocol individualisation may necessitate a refinement in the responder test for other patient subsets. For patients with systemic-onset JRA, additional assessment of fever, extra-articular manifestations, and thrombocytosis/leucocytosis may be useful coprimary endpoints (25). Outcome variables need to be appropriate and consistent with the type of agent under investigation. Investigators should specify, before the trial is initiated, the amount of change that is considered to be clinically important for each outcome variable. Trials should generally last at least six months, except when six-month efficacy data exist in adult RA and there are no reasons to expect loss of efficacy over time. Under these circumstances, trial durations may be blinded/randomised three-month periods; however, six-month periods of open safety data should be obtained. As with adult RA, a three-month trial duration is suggested for NSAIDs.

#### III.2. Prevention of disability

This claim is proposed to reflect durable improvement in physical function and disability in studies of one- to two-year periods with demonstrated improvement in signs and symptoms over the same period. Instruments currently validated for use in JRA include the Childhood Health Assessment Questionnaire (CHAQ), the Juvenile Arthritis Self-Report Index (JASI), and the Juvenile Arthritis Functional Assessment Report (JAFAR). HR-QOL should also be measured and shown not to worsen over the trial.
duration. Endpoints should be tailored to subtypes enrolled in trials (e.g., to assess knee function in pauci-articular JRA patients in whom knee arthritis may be the primary arthritic manifestation). Instruments should be developmentally validated for the age ranges studied in a trial (26).

IV. SUGGESTED READINGS


V. REFERENCES


### Appendix 1. Disability Index of the Health Assessment Questionnaire

Please check the one response which best describes your usual abilities **OVER THE PAST WEEK:**

<table>
<thead>
<tr>
<th>Activity</th>
<th>Without ANY Difficulty</th>
<th>With SOME Difficulty</th>
<th>With MUCH Difficulty</th>
<th>UNABLE To Do</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>DRESSING &amp; GROOMING</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Are you able to:</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>- Dress yourself, including tying shoelaces and doing buttons?</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>- Shampoo your hair?</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>ARISING</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Are you able to:</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>- Stand up from an armless straight chair?</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>- Get in and out of bed?</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>EATING</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Are you able to:</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>- Cut your meat?</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>- Lift a full cup or glass to your mouth?</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>- Open a new milk carton?</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>WALKING</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Are you able to:</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>- Walk outdoors on flat ground?</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>- Climb up five steps?</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Please check any **AIDS** or **DEVICES** that you usually use for any of these activities:

<table>
<thead>
<tr>
<th>Device</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cane</td>
<td>Devices used for dressing (button hook, zipper pull, long-handled shoe horn, etc.)</td>
</tr>
<tr>
<td>Walker</td>
<td>Built up or special utensils</td>
</tr>
<tr>
<td>Crutches</td>
<td>Built up or special chair</td>
</tr>
<tr>
<td>Wheelchair</td>
<td>Other (specify: __________________________)</td>
</tr>
</tbody>
</table>

Please check any categories for which you usually need **HELP FROM ANOTHER PERSON:**

<table>
<thead>
<tr>
<th>Category</th>
<th>Assistance Required</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dressing and Grooming</td>
<td>Eating</td>
</tr>
<tr>
<td>Arising</td>
<td>Walking</td>
</tr>
</tbody>
</table>

331
Please check the one response which best describes your usual abilities **OVER THE PAST WEEK:**

<table>
<thead>
<tr>
<th></th>
<th>Without ANY Difficulty</th>
<th>With SOME Difficulty</th>
<th>With MUCH Difficulty</th>
<th>UNABLE To Do</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>HYGIENE</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Are you able to:</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>- Wash and dry your entire body?</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>- Take a tub bath?</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>- Get on and off the toilet?</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>REACH</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Are you able to:</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>- Reach and get down a 5 pound object (such as a bag of sugar) from just above your head?</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>- Bend down to pick up clothing from the floor?</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>GRIP</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Are you able to:</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>- Open car doors?</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>- Open jars which were previously opened?</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>- Turn faucets on and off?</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>ACTIVITIES</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Are you able to:</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>- Run errands and shop?</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>- Get in and out of a car?</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>- Do chores such as vacuuming or yard work?</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Please check any AID* or DEVICES that you usually use for any of these activities:

<table>
<thead>
<tr>
<th>Raised Toilet Seat</th>
<th>Jar Opener (for jars previously opened)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bathtub Seat</td>
<td>Long-Handled Appliances for Reach</td>
</tr>
<tr>
<td>Bathtub bar</td>
<td>Long-Handled Appliances in Bathroom</td>
</tr>
<tr>
<td></td>
<td>Other (specify: ______________________)</td>
</tr>
</tbody>
</table>

Please check any categories for which you usually need HELP FROM ANOTHER PERSON:

<table>
<thead>
<tr>
<th>Hygiene</th>
<th>Gripping and Opening Things</th>
</tr>
</thead>
<tbody>
<tr>
<td>Reach</td>
<td>Errands and Chores</td>
</tr>
</tbody>
</table>
Appendix 2: SF-36™ Health Status Survey

Patient Identification # ________________
Initials: _____/___/____
Date: _____/___/____

SHORT-FORM HEALTH SURVEY WORKSHEET

INSTRUCTIONS: This survey asks for your views about your health. This information will help keep track of how you feel and how well you are able to do your usual activities.

Answer every question by marking the answer as indicated. If you are unsure about how to answer a question, please give the best answer you can.

1. In general, would you say your health is:
   (circle one)
   Excellent ............................................................................................................................. 1
   Very good ...................................................................................................................... ..... 2
   Good .......................................................................................................................... ..... 3
   Fair .......................................................................................................................... ..... 4
   Poor ......................................................................................................................5

2. Compared to one year ago, how would you rate your health in general now?
   (circle one)
   Much better now than one year ago ................................................................. 1
   Somewhat better now than one year ago .............................................................. 2
   About the same as one year ago ............................................................................ 3
   Somewhat worse now than one year ago ............................................................. 4
   Much worse now than one year ago ................................................................. 5

3. The following items are about activities you might do during a typical day. Does your health now limit you in these activities? If so, how much?
   (circle one number on each line)

<table>
<thead>
<tr>
<th>ACTIVITIES</th>
<th>Yes, Limited A Lot</th>
<th>Yes, Limited A Little</th>
<th>No, Not Limited At All</th>
</tr>
</thead>
<tbody>
<tr>
<td>a. Vigorous activities, such as running, lifting heavy objects, participating in strenuous sports</td>
<td>1</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>b. Moderate activities, such as moving a table, pushing a vacuum cleaner, bowling, or playing golf</td>
<td>1</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>c. Lifting or carrying groceries</td>
<td>1</td>
<td>2</td>
<td>3</td>
</tr>
</tbody>
</table>
d. Climbing *several* flights of stairs  & 1 & 2 & 3  

e. Climbing *one* flight of stairs  & 1 & 2 & 3  

f. Bending, kneeling, or stooping  & 1 & 2 & 3  

g. Walking *more than one mile*  & 1 & 2 & 3  

h. Walking *several blocks*  & 1 & 2 & 3  

i. Walking *one block*  & 1 & 2 & 3  

j. Bathing or dressing yourself  & 1 & 2 & 3  

4. During the past 4 weeks, have you had any of the following problems with your work or other regular daily activities as a result of your physical health?

(circle one number on each line)

<table>
<thead>
<tr>
<th></th>
<th>YES</th>
<th>NO</th>
</tr>
</thead>
<tbody>
<tr>
<td>a. Cut down on the <em>amount of time</em> you spent on work or other activities</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>b. <em>Accomplished less</em> than you would like</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>c. Were limited in the <em>kind</em> of work or other activities</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>d. Had difficulty performing the work or other activities (for example, it took extra effort)</td>
<td>1</td>
<td>2</td>
</tr>
</tbody>
</table>

5. During the past 4 weeks, have you had any of the following problems with your work or other regular daily activities as a result of any emotional problems (such as feeling depressed or anxious)?

(circle one number on each line)

<table>
<thead>
<tr>
<th></th>
<th>YES</th>
<th>NO</th>
</tr>
</thead>
<tbody>
<tr>
<td>a. Cut down on the <em>amount of time</em> you spent on work or other activities</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>b. <em>Accomplished less</em> than you would like</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>c. Didn’t do work or other activities as <em>carefully</em> as usual</td>
<td>1</td>
<td>2</td>
</tr>
</tbody>
</table>

6. During the past 4 weeks, to what extent has your physical health or emotional problems interfered with your normal social activities with family, friends, neighbors, or groups?

(circle one)

Not at all..................................................................................................................... 1  
Slightly ...................................................................................................................... 2  
Moderately................................................................................................................... 3  
Quite a bit.................................................................................................................. 4  
Extremely................................................................................................................... 5  

334
7. How much bodily pain have you had during the past 4 weeks?

(circle one)

None .......................................................................................................................... ..... 1
Very mild ....................................................................................................................... 2
Mild ............................................................................................................................. ..... 3
Moderate ......................................................................................................................... 4
Severe ........................................................................................................................... ....... 5
Very severe ..................................................................................................................... 6

8. During the past 4 weeks how much did pain interfere with your normal work (including both work outside the home and housework)?

(circle one)

Not at all..................................................................................................................... ......... 1
A little bit ............................................................................................................................ 2
Moderately .......................................................................................................................... 3
Quite a bit .......................................................................................................................... 4
Extremely ........................................................................................................................ 5

9. These questions are about how you feel and how things have been with you during the past 4 weeks. For each question, please give one answer that comes closest to the way you have been feeling. How much of the time during the past 4 weeks----

(circle one number on each line)

<table>
<thead>
<tr>
<th>ACTIVITIES</th>
<th>All of the Time</th>
<th>Most of the Time</th>
<th>A Good Bit of the Time</th>
<th>Some of the Time</th>
<th>A Little of the Time</th>
<th>None of the Time</th>
</tr>
</thead>
<tbody>
<tr>
<td>a. Did you feel full of pep?</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>6</td>
</tr>
<tr>
<td>b. Have you been a very nervous person?</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>6</td>
</tr>
<tr>
<td>c. Have you felt so down in the dumps that nothing could cheer you up?</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>6</td>
</tr>
<tr>
<td>d. Have you felt calm and peaceful?</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>6</td>
</tr>
<tr>
<td>e. Did you have a lot of energy?</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>6</td>
</tr>
<tr>
<td>f. Have you felt downhearted and blue?</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>6</td>
</tr>
<tr>
<td>g. Did you feel worn out?</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>6</td>
</tr>
<tr>
<td>h. Have you been a happy person?</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>6</td>
</tr>
<tr>
<td>i. Did you feel tired?</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>6</td>
</tr>
</tbody>
</table>
10. During the past 4 weeks how much did pain interfere with your normal work (including both work outside the home and housework)?

   (circle one)

   Not at all..................................................................................................................... 1
   A little bit ....................................................................................................................... 2
   Moderately ..................................................................................................................... 3
   Quite a bit ....................................................................................................................... 4
   Extremely ....................................................................................................................... 5

11. These questions are about how you feel and how things have been with you during the past 4 weeks. For each question, please give one answer that comes closest to the way you have been feeling. How much of the time during the past 4 weeks----

   (circle one number on each line)

12. During the past 4 weeks, how much of the time has your physical health or emotional problems interfered with your social activities (like visiting your friends, relatives, etc.)?

   (circle one)

   All of the time ....................................................................................................... 1
   Most of the time ................................................................................................... 2
   Some of the time .................................................................................................. 3
   A little of the time ............................................................................................... 4
   None of the time .................................................................................................. 5

13. How TRUE or FALSE is each of the following statements for you?

   (circle one number on each line)

<table>
<thead>
<tr>
<th>ACTIVITIES</th>
<th>Definitely True</th>
<th>Mostly True</th>
<th>Don't Know</th>
<th>Mostly False</th>
<th>Definitely False</th>
</tr>
</thead>
<tbody>
<tr>
<td>a. I seem to get sick a little easier than other people</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
</tr>
<tr>
<td>b. I am as healthy as anybody I know</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
</tr>
<tr>
<td>c. I expect my health to get worse</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
</tr>
<tr>
<td>d. My health is excellent</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
</tr>
</tbody>
</table>
Pharmacological Research in Other Disorders

Chapter 28. Neoplastic Diseases

Chapter 29. Analgesic Drugs for Cancer Pain Management

Chapter 30. Osteoporosis
Chapter 28. Neoplastic Diseases

Wendy Parulekar, M.D., FRCP(C) Med Onc

Clinical Trials Group
National Cancer Institute of Canada
10 Stuart Street
Queen's University
Kingston, Ontario K7L 3N6
CANADA
I. INTRODUCTORY REMARKS

Cancer is a global health problem. While there have been some major advances in the treatment of solid tumours over the past 30 years, most therapeutic successes have been relatively modest, leading to survival gains of a few percentage points at best. Thus there is an urgent need for continued clinical research involving new drugs or drug combinations. In acknowledgement of the intense resource and ethical implications of exposing patients with life threatening illnesses to experimental and potentially ineffective therapies, there exists an equally compelling need for well designed and conducted studies.

Chemotherapy is the mainstay of drug therapy for most solid tumours. These agents generally target DNA or the mitotic apparatus. In general, the principles behind chemotherapy dosing are based on preclinical data demonstrating that there is a direct relationship between dose and tumour cell kill as well as dose and toxicity. Thus selection of the appropriate dose for treatment represents a balance between antitumour effect and side effects.

The initial evaluation of new agents in human subjects occurs within the context of phase I and II studies. Phase I or dose finding studies seek to determine the appropriate dose for further study (recommended phase II dose or RPTD). This typically involves exposure of successive cohorts of 3 to 6 patients with various tumour types to increasing doses of drug. Careful evaluation of the toxicity and pharmacokinetic profiles of the new agent(s) occurs at each dose level. The RPTD is defined as that dose which produces serious but reversible side effects in a predefined proportion of patients. Subjects with advanced, heavily pretreated, disease are usually included in phase I trials provided that they have adequate organ and functional status as defined within the protocol.

Phase II studies screen for activity of a new drug or drug combination using the RPTD determined in the phase I study. Previously untreated or minimally treated patients with susceptible tumour types based on preclinical and early clinical evidence are included. The primary endpoint is estimation of the objective response rate, which is usually defined as the proportion of patients who have partial or complete shrinkage of tumour after drug exposure according to predefined standard criteria (1). In addition to characterization of the toxicity profile of the new therapy, these studies may also include pharmacokinetic or other pharmacodynamic endpoints. Many approaches exist for the conduct of phase II studies in cancer medicine including those that incorporate progression (2) and toxicity (3) information as part of the criteria for early termination of a trial. A common approach is that of the two stage design (4): if a minimum predetermined number of responses are seen with the first cohort of patients then the accrual is continued to the second stage to provide a more reliable estimation of activity. Sample size calculation is performed by specifying a response rate of interest as well as a lower response rate level below which the drug will be declared inactive. Traditional phase II design is non-comparative although randomization may be used to improve the efficiency of this type of study as a screening tool (5).

Phase III studies serve as the definitive tests of efficacy of new therapies. Using a randomized design to minimize bias, patients are allocated to the new agent(s) of interest or the standard therapy. In the field of oncology, these studies are powered to detect clinically meaningful differences in relevant endpoints such as overall or disease free survival. Quality of life or palliation of disease related symptoms may be the primary objective of symptom control studies. Phase III trials are generally resource intense due to the large sample sizes and duration of follow up generally required. Careful consideration must therefore be given to the primary study question and design since it is unethical to involve patients and investigators in trials addressing clinically irrelevant issues or involving poor methodology.
II. PHASE II STUDIES FOR REGISTRATION OF NEW THERAPIES

II.1. Outline of a typical development plan
Patients with advanced tumours of a single specific histological type are enrolled in a single arm, non-comparative study of a new drug/drug combination. The study may be conducted in a single or multi-institutional fashion although the latter is preferred to better estimate general feasibility of delivery of the new drug regimen.

II.2. Study plan

II.2.A. Study Objectives
Primary objectives:
   a. To estimate the activity of drug X given in schedule Y (mg/route/frequency) in patients with previously untreated advanced tumours of a particular histology.

Secondary objectives:
   a. To assess the toxic effects (or “adverse effects”) of drug X in patients with previously untreated advanced tumours of a particular histology.
   b. In some studies consider also: To describe the relationship between molecular tumour characteristics and objective response.

II.2.B. Primary Endpoints
   a. Objective tumour response for solid tumours is assessed using the RECIST criteria.\textsuperscript{1} In phase II studies of new agents in hematological malignancies, response may be measured using peripheral blood indices (hemoglobin, white blood cell count, platelets, presence of malignant cells), bone marrow (cellularity and % of malignant cells) and cytogenetics.

II.2.C. Secondary Endpoints
   a. Duration of response.
   b. Adverse effects (toxic effects) in patients receiving drug X given in schedule Y as categorized and graded using the Common Terminology Criteria for Adverse Events (CTCAE) Version 3 (6).

II.2.D. Exploratory Endpoints
   a. Molecular measure of drug effect in tumour or surrogate tissue.

II.2.E. Study Design
Patients will be entered on this multi institutional, open label, single arm cohort study provided that all eligibility criteria are met and informed consent has been obtained.

Treatment must begin within 2 days of study entry. All patients will begin treatment at the protocol mandated dose and dose adjustments will be made on the basis of adverse (toxic) effects as required. Response and adverse event evaluation will be measured according to standard criteria as noted above.

All concomitant therapy, including alternative therapies, must be recorded on the case report forms. Other cytotoxic chemotherapy or investigational anti-cancer agents are not permitted.

II.2.F. Planned Sample Size
A typical sample size calculation will employ the Simon Two Stage Phase II Design method. Utilizing a response probability of interest (Ha) of 30%, a minimal response probability of 10% (Ho), error probabilities of 5% for accepting the drug with the minimal response probability and 20% for rejecting the drug with the response probability of interest:
Stage I of accrual: 10 response evaluable patients will be entered in the first stage. Using Ho ≤ 10% and Ha ≥ 30%, the drug will be declared inactive at the end of the first stage if there are fewer than 2 objective responses.

Stage II of accrual: If the above criterion is not met then 19 additional patients will be accrued onto the study for a final sample size of 29. The drug will be declared active if there are greater than 5 objective responses in the total sample.

II.2.G. Study Population
Previously untreated or minimally treated patients with susceptible tumour types based on pre clinical and early clinical evidence are included.

II.2.H. Specific inclusion criteria
Patients will be considered eligible for study entry provided that the following criteria are met:
   a. Histologically documented advanced/recurrent solid tumour of the specific histological type under evaluation.
   b. Presence of clinically or radiologically documented disease. At least one site of disease must be unidimensionally measurable defined as follows:
      • X – ray, physical exam ≥ 20mm
      • Spiral CT scan ≥ 10 mm
      • Non-spiral CT ≥ 20 mm
   c. Patients must have a life expectancy of at least 12 weeks.
   d. Age ≥ 18 years.
   e. Eastern Cooperative Oncology Group (ECOG) performance status of 0, 1, or 2 (See Appendix 1).
   f. Previous therapy: Prior adjuvant therapy is permitted but patients must not have had systemic therapy for advanced/recurrent disease.
      • Patients must be at least ≥ 6 months since the last dose of adjuvant chemotherapy, if applicable.
      • Patients may have received prior radiation provided that all of the following conditions are met:
         i. There is measurable disease outside the previously irradiated area. Patients whose sole site of disease is in a previously irradiated area are ineligible unless there is evidence of progression or new lesions in the irradiated field.
         ii. At least 4 weeks must have lapsed since the last treatment with radiation.
         iii. Surgery is permissible provided that at least 4 weeks have lapsed since any major surgery.
   g. Laboratory Requirements: (must be within 7 days prior to study entry)
      • Hematology:
         i. Absolute granulocyte count (AGC) ≥ 1.5 x 10^9/L
         ii. Platelets ≥ 100 x 10^9/L
      • Chemistry:
         i. Serum creatinine ≤ Upper Normal Limit
         ii. Bilirubin ≤ Upper Normal Limit
         iii. AST ≤ 2.5 x Upper Normal Limit
   h. Patient consent must be obtained according to local institutional policy of University Human Experimentation Committee requirements.
   i. Patients must be accessible for treatment and follow-up.

II.2.I. Specific Exclusion Criteria
   a. Prior history of other malignancies except: adequately treated non-melanoma skin cancer, curatively treated in-situ cancer of the cervix or other solid tumours curatively treated with no evidence of disease for ≥ 5 years.
b. Prior chemotherapy for advanced/recurrent disease.
c. Non-measurable disease only.
d. Uncontrolled intercurrent illness including, but not limited to, ongoing or active infection, symptomatic congestive heart failure, unstable angina pectoris, cardiac arrhythmias or psychiatric illness/ social situation that would limit study compliance.
e. Concurrent treatment with anticancer or investigational agents including hormonal therapy.
f. Symptomatic brain metastases.
g. History of allergic reactions attributed to compounds of similar chemical or biologic composition to the study drug.
h. Pregnant or lactating women. All patients of child bearing potential must use adequate contraception while on study.

II.2.J. Tools for Assessment of Endpoints

Objective Response Criteria
Response and progression will be evaluated in this study using the RECIST criteria. Changes in only the largest diameter (unidimensional measurement) of the tumour lesions are used in the RECIST criteria.

Measurable Disease
Measurable lesions are defined as those that can be accurately measured in at least one dimension (longest diameter to be recorded) as $\geq 20$ mm with conventional techniques (physical examination, CT, x-ray, MRI) or as $\geq 10$ mm with spiral CT scan. All tumour measurements must be recorded in millimetres (or decimal fractions of centimetres).

Non-measurable Disease
All other lesions (or sites of disease), including small lesions (longest diameter $< 20$ mm with conventional techniques or $< 10$ mm with spiral CT scan) are considered non-measurable disease. Bone lesions, leptomeningeal disease, ascites, pleural/pericardial effusions, lymphangitis cutis/pulmonis, inflammatory breast disease, abdominal masses (not followed by CT or MRI) and cystic lesions are all non-measurable.

Target Lesions
All measurable lesions up to a maximum of 5 lesions per organ and 10 lesions in total representative of all involved organs should be identified as target lesions and be recorded and measured at baseline. Target lesions should be selected on the basis of their size (lesions with the longest diameter) and their suitability for accurate repetitive measurements (either by imaging techniques or clinically). A sum of the longest diameter (LD) for all target lesions will be calculated and reported as the baseline sum LD. The baseline sum LD will be used as reference to further characterize the objective tumour response of the measurable dimension of the disease. If there are $> 10$ measurable lesions, those not selected as target lesions will be considered together with non-measurable disease as non-target lesions.

Non-target Lesions
All non-measurable lesions (or sites of disease) plus any measurable lesions over and above the 10 listed as target lesions. Measurements are not required but these lesions should be noted at baseline and should be followed as “present” or “absent”.

All patients who receive at least one cycle of therapy and have their disease re-evaluated will be considered evaluable for response. All patients will have their BEST RESPONSE on study classified as outlined below:

Complete Response (CR): disappearance of all clinical and radiological evidence of tumour (both target and non-target).
Partial Response (PR): at least a 30% decrease in the sum of LD of target lesions, taking as reference the baseline sum LD.

Stable Disease (SD): steady state of disease. Neither sufficient shrinkage to qualify for PR nor sufficient increase to qualify for PD.

Progressive Disease (PD): at least a 20% increase in the sum of LD of measured lesions taking as references the smallest sum LD recorded since the treatment started. Appearance of new lesions will also constitute progressive disease. In exceptional circumstances, unequivocal progression of non-target lesions may be accepted as evidence of disease progression.

<table>
<thead>
<tr>
<th>Target Lesions</th>
<th>Non-Target Lesions</th>
<th>New Lesions</th>
<th>Overall Response</th>
<th>Best Response for this category also requires</th>
</tr>
</thead>
<tbody>
<tr>
<td>CR</td>
<td>CR</td>
<td>No</td>
<td>CR</td>
<td>≥ 4 wks. confirmation</td>
</tr>
<tr>
<td>CR</td>
<td>Non-CR/Non-PD</td>
<td>No</td>
<td>PR</td>
<td>≥ 4 wks. confirmation</td>
</tr>
<tr>
<td>PR</td>
<td>Non-PD</td>
<td>No</td>
<td>PR</td>
<td>documented at least once ≥ 4 wks. from baseline</td>
</tr>
<tr>
<td>SD</td>
<td>Non-PD</td>
<td>No</td>
<td>SD</td>
<td>no prior SD, PR or CR</td>
</tr>
<tr>
<td>PD</td>
<td>Any</td>
<td>Yes or No</td>
<td>PD</td>
<td></td>
</tr>
<tr>
<td>Any</td>
<td>PD*</td>
<td>Yes or No</td>
<td>PD</td>
<td></td>
</tr>
<tr>
<td>Any</td>
<td>Any</td>
<td>Yes</td>
<td>PD</td>
<td></td>
</tr>
</tbody>
</table>

* In exceptional circumstances, unequivocal progression in non-target lesions may be accepted as disease progression.

Note: Patients with a global deterioration of health status requiring discontinuation of treatment without objective evidence of disease progression at that time should be reported as “symptomatic deterioration”. Every effort should be made to document the objective progression even after discontinuation of treatment.

Quality of life is measured using validated instruments such as the EORTC QLQ C30 (7). Patient diaries are appropriate for symptom control studies.

Specific Criteria for Early Withdrawal and Discontinuation: See Phase II Section.

Tools for Assessment of Primary Endpoints:
The same method of assessment and the same technique should be used to characterize each identified and reported lesion at baseline and during follow-up.

**Clinical Lesions.** Clinical lesions will only be considered measurable when they are superficial (e.g. skin nodules, palpable lymph nodes). For the case of skin lesions, documentation by colour photography including a ruler to estimate the size of the lesion is recommended.

**Chest X-ray.** Lesions on chest X-ray are acceptable as measurable lesions when they are clearly defined and surrounded by aerated lung. However, CT is preferable.

**CT, MRI.** CT and MRI might be the best currently available and reproducible methods to measure target lesions selected for response assessment. Conventional CT and MRI should be performed with cuts of 10 mm or less in slice thickness contiguously. Spiral CT should be performed using a 5 mm contiguous reconstruction algorithm. This applies to the chest, abdomen and pelvis. Head & neck and extremities usually require specific protocols.
Ultrasound. When the primary endpoint of the study is objective response evaluation, ultrasound (US) should not be used to measure tumour lesions that are clinically not easily accessible. It is a possible alternative to clinical measurements for superficial palpable nodes, subcutaneous lesions and thyroid nodules. US might also be useful to confirm the complete disappearance of superficial lesions usually assessed by clinical examination.

Endoscopy, Laparoscopy. The utilization of these techniques for objective tumour evaluation has not yet been fully and widely validated. Their uses in this specific context require sophisticated equipment and a high level of expertise that may only be available in some centres. Therefore, the utilization of such techniques for objective tumour response should be restricted to validation purposes in reference centres. However, such techniques can be useful to confirm complete pathological response when biopsies are obtained.

Tumour Markers. Tumour markers alone cannot be used to assess response. If markers are initially above the upper normal limit, they must normalize for a patient to be considered in complete clinical response. Specific additional criteria for standardized usage of PSA and CA-125 response in support of clinical trials are being developed.

Cytology, Histology. These techniques can be used to differentiate between PR and CR in rare cases (for example, residual lesions in tumour types such as germ cell tumours, where known residual benign tumours can remain).

The cytological confirmation of the neoplastic origin of any effusion that appears or worsens during treatment when the measurable tumour has met criteria for response or stable disease is mandatory to differentiate between response or stable disease (an effusion may be a side effect of the treatment) and progressive disease.

Tools for Assessment of Secondary Endpoints:
Response Duration. Response duration will be measured from the time measurement criteria for CR/PR (whichever is first recorded) are first met until the first date that recurrent or progressive disease is objectively documented.

Stable Disease Duration. Stable disease duration will be measured from the time of start of therapy until the criteria for progression are met, taking as reference the smallest measurements recorded since the treatment started.

Adverse Events. All patients will be evaluable for assessment of adverse events from the time of their first dose of study drug.

Adverse events will be monitored on an ongoing basis by the study principal investigator and, if applicable, the study coordinating office. Adverse events will be categorized using the CTCAE Version 3.0. The worst event for each patient in each category or subcategory will be described. Events related and unrelated to treatment will be captured.

II.2.K. Specific Criteria for Early Withdrawal and Discontinuation
Patients may stop protocol treatment in the following instances:
  a. Intercurrent illness which would, in the judgement of the investigator, affect assessments of clinical status to a significant degree and require discontinuation of protocol therapy.
  b. Unacceptable toxicity: Patients with intolerable or limiting toxicity despite dose reductions as defined in the protocol may be removed from study as well as those patients with Grade 3 or 4 toxicities which do not improve to \( \leq \) Grade 1 despite drug hold for 2 weeks.
c. Tumour progression as defined by RECIST criteria.

d. Patient request.

II.2.L. Data Analysis Method

Enrollment will occur using a 2-stage design described above. Using the null hypothesis that the response rate is 10% and the alternate hypothesis that the response rate is 30%, the sample size will yield a significance level of 5% and a power of 80%.

Primary Endpoints:
Using the primary endpoint of overall survival in study of patients with advanced disease as an example, the data analysis will involve the generation of survival curves for each treatment arm. All randomized patients will be included in the primary analysis. Survival will be defined as the time from randomization to death from any cause. Patients who are alive at the time of final analysis will be censored at the time of last contact. A Kaplan-Meier curve for proportions of survival in each treatment arm will be displayed and 95% confidence intervals for median survival computed using the method of Brookmeyer and Crowley (8). The two treatment arms will be compared using the log-rank test adjusted for the stratification variables. In addition, the effect of study centre and other potential prognostic factors on overall survival will be assessed using Cox regression. The Schoenfeld residual plots will be used to check the model assumption for the Cox regression (9).

Secondary Endpoints:
Progression free survival (PFS):
This is defined as the time from randomization to the first observation of disease progression according to the RECIST criteria or death due to any cause. A patient who stops treatment with study drug and receives alternative therapy prior to documentation of disease progression will be censored on the date that the alternative therapy begins. If a patient has not progressed or received alternative therapy, PFS will be censored on the date of the last disease assessment. All analyses for overall survival will be similarly performed for PFS.

Response Rate:
Patients will be evaluable for objective tumour response if they have at least one measurable lesion at baseline and at least one disease assessment after baseline. In addition, patients who develop PD prior to this time will also be considered evaluable for response. The response rate will be estimated as the proportion of patients evaluable for response who meet the criteria for complete or partial response. A Cochran-Mantel-Haenszel test will be used to compare tumour response rate between arms adjusting for stratification factors.

The duration of response will be calculated for all patients achieving a PR or CR. Duration of PR/CR is defined as the time from first objective status assessment of CR/PR to the first time disease progression or death is documented. A patient who stops treatment with all study drugs and receives alternative therapy prior to documentation of disease progression, will be censored on the date alternative therapy begins. The date of progression will be considered as the event date for the duration of response. If a patient has not progressed or died, the duration of response will be censored on the date of the last known alive date. The duration of response will be analyzed using similar methods described for overall survival.

Quality of Life:
The quality life of patients will be assessed using EORTC QLQ-C30. The EORTC QLQ-C30 is a validated and reliable self-administered cancer specific questionnaire with multi-dimensional scales. It consists of both multi-item scales and single item measures, including five functioning domains, a global quality of life domain, three symptom domains and six single items. For each domain or single item measure a linear transformation will be applied to standardize the raw score to range between 0 and 100.
Since quality of life will be assessed longitudinally, the method of analysis of variance for repeated measures will be used for domains represented by aggregate scores.

Toxicity:
All patients who receive at least one dose of protocol therapy will be included in the safety analysis. Descriptive summary tables will be presented on safety parameters by treatment arm. Toxicity rates will be compared between treatment arms using the Fisher’s Exact Test, as needed. Oversight of the study by an independent data safety monitoring committee (DSMC) will occur. Included in the mandate of this committee will be ongoing review of the toxicity experience on trial and any interim analysis results as specified in the protocol (10).

III. PHASE III STUDIES FOR REGISTRATION OF NEW THERAPIES

III.1. Outline of a typical development plan

A promising drug/drug combination is selected for further study based on preclinical and early clinical evidence. Using an active and tolerable dose of therapy as defined in phase I and II studies, the relevant patient population (histological subtype, disease stage) and study question are chosen for the phase III trial.

III.2. Study plan

III.2.A. Study Objectives
As for Phase II studies.

III.2.B. Primary endpoints
Overall survival

III.2.C. Secondary endpoints
a. Disease free survival (adjuvant trials).
b. Progression free survival (advanced disease trials).
c. Toxicity.
d. Response rate and duration of response (advanced disease setting).
e. Quality of life.

III.2.D. Exploratory endpoints
a. Relationship between molecular characteristics of tumour and prognosis.
b. Relationship between molecular characteristics of tumour and probability of response to therapy.

III.2.E. Study Design
A randomized parallel group design is used. Blinding of treatment assignments may be appropriate, particularly in studies involving quality of life or symptom control endpoints.

Prior therapy may be allowed depending on the disease under evaluation but, in general, previously untreated or minimally treated patients with good functional status are assigned to the experimental or control arm as defined by the protocol.

Treatment: For advanced disease studies, therapy is generally continued until progression or occurrence of dose limiting toxicity. In early disease or adjuvant studies, protocol therapy is given for a fixed number of cycles or duration.
Treatment must begin within 2 days of study entry. All patients will begin treatment at the protocol mandated dose and dose adjustments will be made on the basis of adverse (toxic) effects as required. Response and adverse event evaluation will be measured according to standard criteria as noted above.

All concomitant therapy, including alternative therapies, must be recorded on the case report forms. Other cytotoxic chemotherapy or investigational anti-cancer agents are not permitted.

III.2.F. Planned Sample Size
Sample size calculations for clinical trials require specification of the type I and II errors as well as the magnitude of difference in outcome that the trial is designed to detect. The latter specification is a clinical one and depends on what difference in efficacy is likely to be present between the treatments and what difference would change current practice if detected.

As an example, a study in the advanced disease setting is designed to compare overall survival between patients randomized to ARM 1 (control) and patients randomized to ARM 2 (experimental). Based on other clinical data, the median survival of patients randomized to ARM 1 is estimated to be 0.55 years. In order to have 80% power to detect a 33% improvement in median survival in the experimental arm (hazard ratio of 1.33) using a two-sided 5% significance test, 381 deaths must be observed before the final analysis. In anticipation of accrual of 450 patients in 9 months, the required number of deaths (381) would be observed after following all patients for another 18 months.

Some studies employ interim analyses with adjusted p values to detect early and potentially clinically significant differences in outcome between arms before the final analysis. If interim analyses are planned, this should be clearly indicated in the body of the protocol.

Balance between the treatment arms for known prognostic factors is achieved by use of stratification factors at the time of randomization. Examples of typical stratification factors in oncology include disease stage, sex and performance status.

III.2.G. Study Population
Using a randomized design to minimize bias, patients are allocated to the new agent(s) of interest or the standard therapy.

III.2.H. Specific inclusion criteria
Inclusion Criteria: See Phase II section for typical eligibility criteria.

For advanced studies, prior systemic therapy may be allowed as long as an appropriate time lapse has occurred at the time of randomization. For adjuvant studies, no prior systemic therapy is generally allowed.

Patients with measurable or unmeasurable disease may be appropriate for advanced disease studies in which survival is the primary endpoint. Adjuvant study protocols for patients with early disease mandate the absence of tumour (clinically and radiologically) at the time of randomization.

III.2.I. Specific Exclusion Criteria
See Phase II section for typical ineligibility criteria

III.2.J. Tools for Assessment of Endpoints
Tools for Assessing Primary Endpoints:
For overall survival, the date of death is determined from hospital records or the death certificate whenever possible.
Tools for Assessing Secondary Endpoints:
In early disease studies measuring disease free survival, the date of relapse is the first date of clinical or radiological relapse.

Recurrence will be categorized as local, regional or distant, based on the histology and location of the primary tumour. The date of first recurrence should always be based on the onset of a sign of recurrence rather than the onset of a symptom.

The date of first detection of a palpable lesion is acceptable only when the diagnosis of tumour involvement is subsequently established.

The diagnosis of recurrent disease by radiographs or scans should be dated from the date of the first positive record, even if this is determined in retrospect.

For toxicity and response assessment, see Phase II Section regarding use of appropriate tools.

Quality of life is measured using validated instruments such as the EORTC QLQ C30 (7). Patient diaries are appropriate for symptom control studies.

III.2.K. Specific Criteria for Early Withdrawal and Discontinuation
See Phase II Section

III.2.L. Methods of Measurement
See Phase II Section

IV. EXAMPLES OF LANDMARK WELL DESIGNED TRIALS

IV.1. Phase II studies


IV.2. Phase III studies


V. SUGGESTED READINGS

V.1. Phase II studies

V.2. Phase III studies

VI. REFERENCES

# APPENDIX I - PERFORMANCE STATUS (ECOG)

<table>
<thead>
<tr>
<th>Grade</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>Fully active, able to carry on all pre-disease performance without restriction (Karnofsky 90-100).</td>
</tr>
<tr>
<td>1</td>
<td>Restricted in physically strenuous activity but ambulatory and able to carry out work of a light or sedentary nature (eg, light housework, office work) (Karnofsky 70-80).</td>
</tr>
<tr>
<td>2</td>
<td>Ambulatory and capable of all self-care but unable to carry out any work activities. Up and about more than 50% of waking hours (Karnofsky 50-60).</td>
</tr>
<tr>
<td>2</td>
<td>Capable of only limited self-care, confined to bed or chair more than 50% of waking hours (Karnofsky 30-40).</td>
</tr>
<tr>
<td>3</td>
<td>Completely disabled. Cannot carry on any self-care. Totally confined to bed or chair (Karnofsky 10-20).</td>
</tr>
</tbody>
</table>
Chapter 29. Analgesic Drugs for Cancer Pain Management

Kim Fisher, Ph.D. 1,2
Neil Hagen, M.D. 1,3
Patrick du Souich, M.D., Ph.D. 4
John Oldenhof, Ph.D. 5
Myroslava Romach, M.D. 5
Anh Ho Ngoc, Ph.D. 6
Edward Sellers, M.D., Ph.D. 5,7

1 Alberta Cancer Board Palliative Care Research Initiative
2 Department of Obstetrics and Gynecology
   University of Calgary
   Calgary, Alberta
   CANADA
3 Division of Palliative Medicine
   Department of Oncology, and Departments of Clinical Neurosciences and Medicine
   University of Calgary
   Calgary, Alberta
   CANADA
4 Département de pharmacologie
   Faculté de médecine
   Université de Montréal
   Montréal, Québec H3C 3J7
   CANADA
5 Ventana Clinical Research Corporation
   Toronto, Ontario
   CANADA
6 International Wex Technologies Inc.
   Vancouver, British Columbia
   CANADA
7 Departments of Pharmacology, Medicine and Psychiatry
   University of Toronto
   Toronto, Ontario
   CANADA
I. INTRODUCTORY REMARKS

Pain directly related to cancer or caused by treatments for cancer is a highly prevalent clinical problem. Between 30-85% of patients experience pain at some point during the illness trajectory, with estimates of 18-78% experiencing substantial pain in developed countries.

The current therapies for severe pain in cancer patients remain unsatisfactory. Opioids, the mainstay of cancer pain therapy, have a number of limitations. First, they are accompanied by a significant risk of dose limiting toxicity such as constipation, nausea and drowsiness. Second, some types of cancer pain, such as neuropathic pain and movement related bone pain, are difficult pain problems and carry a less good prognosis for control by opioids at doses that are tolerable. Third, opioids carry a significant perceived risk of abuse potential by society in general, which may at times represent barriers to the effective management of cancer pain. Fourth, cancer patients are at high risk for organ failure such as renal impairment or liver disease as their underlying disease progresses, and this comorbidity increases the risk of multifactorial, dose limiting symptoms associated with opioid administration such as delirium.

For cancer pain problems that have limited responsiveness to opioids, the World Health Organization (WHO) recommends the use of non-opioid (e.g. non-steroidal anti-inflammatory drugs) and adjuvant analgesics, in addition to opioids. Adjuvant analgesics include tricyclic antidepressants, anticonvulsants, N-methyl-D-aspartate antagonists, corticosteroids and others. The likelihood of effectiveness of many of the major classes of analgesics for difficult cancer pain problems remains disputed. There is an urgent need for controlled trials to evaluate the efficacy and adverse effect profiles of currently available non-opioid and adjuvant analgesics for cancer pain. There is also a need for newer analgesic drugs that have greater efficacy for severe cancer pain, better adverse effect profiles, and a reduced potential for toxicity, tolerance and addiction.

This chapter profiles a randomized, double-blind, placebo-controlled parallel-group trial design to evaluate the efficacy and safety of investigational drug XXX for patients with moderate to severe cancer pain. Based on the results from previous clinical studies, it is known that this drug has several important properties. Specifically, drug XXX has a rapid onset of action (5-30 minutes post injection), cumulative analgesic effect with twice-daily injections, long duration of action extending for days or weeks beyond the treatment period, and a favourable adverse effect profile at the studied doses. The trial described in this chapter is designed to better characterize these properties of drug XXX. For specific aspects of clinical trial design, the authors will also broaden the discussion to alternative approaches to help guide the development of Phase II clinical trials for analgesic drugs with different pharmacokinetic and pharmacodynamic properties. The examples in this chapter are most applicable to trials for evaluating analgesic drugs intended to be used in addition to patients’ current analgesic therapy.

II. PHASE II STUDIES FOR REGISTRATION OF NEW ANALGESIC DRUGS

II.1. Outline of a typical development plan

Early Phase II analgesic studies are typically multiple-dose clinical trials designed to identify the optimal dose for later Phase II studies, and to characterize the drug’s analgesic and adverse effect profiles. Once the optimal dose has been identified, later Phase II analgesic studies tend to be multicenter, randomized, double-blind, controlled trials involving larger sample sizes. When evaluating drugs that are intended to be used in addition to opioid therapy, the use of a placebo control is appropriate and ethical. There are two major types of randomized, double-blind, placebo-controlled trial designs in analgesic studies: parallel-group and crossover. However, if there is a significant chance that the drug will have a carryover effect or the duration of a drug’s effect is difficult to predict or is variable, a parallel-group design should be used. One should also consider that a crossover design extends the length of a trial. For the cancer pain
population progression of the disease over the course of many weeks and a consequential worsening of pain over the duration of the trial could complicate the interpretation of the results.

II.2. Short term studies

The following example will be used to illustrate a pivotal, later Phase II analgesic study: a multicentre, randomized, double-blind, placebo-controlled, parallel-group trial to evaluate the efficacy and safety of subcutaneously administered drug XXX in patients with moderate to severe cancer pain. For this example, drug XXX or placebo is administered b.i.d. (twice each day) to patients over a four day period.

II.2.A. Study Objectives

Below is an example of a set of objectives for a clinical trial to evaluate investigational drug XXX as an adjuvant to opioid therapy for patients with moderate to severe cancer pain.

**Primary Objectives**

a. to compare the efficacy of subcutaneous (s.c.) XXX treatment versus placebo in reducing the intensity of pain
b. to compare the safety of s.c. XXX treatment versus placebo

**Secondary Objectives**

a. to estimate the onset of analgesic effect of s.c. XXX
b. to estimate the time of peak analgesic effect of s.c. XXX
c. to estimate the duration of analgesic effect of s.c. XXX treatment
d. to determine whether s.c. XXX treatment reduces the need for breakthrough medication
e. to determine whether s.c. XXX treatment improves patient function.

**Exploratory Objective**

a. to generate preliminary information about the specificity of XXX’s analgesic action for neuropathic, visceral, and somatic pain

II.2.B. Primary endpoints

A wide variety of assessment tools are used for study endpoints to evaluate the efficacy of analgesic drugs in clinical trials and, unfortunately, there is currently no one gold standard for this very important task. Choice of pain intensity measures varies considerably across published clinical trials. These measures include variations in numeric rating scales, visual analogue scales, and verbal rating scales. All three of these scale types have been shown to be sensitive to treatment- and time-related changes in pain intensity. Pain intensity measures are also diverse with respect to whether they measure worst pain, average pain, or least pain over a specified period of time, or current pain. The Brief Pain Inventory (BPI) and other instruments use several pain intensity measures, concurrently, to estimate the degree of this dimension of pain. Measures of pain relief or patient evaluation of analgesic efficacy should also be considered in addition to pain intensity measures, since they can be effective in detecting a clinically meaningful analgesic response to treatment. However, in the literature there is considerable variation in the types of scales used to assess these two constructs. Pain intensity and pain relief measures are often concurrently used to characterize the onset, peak, and duration of an investigational drug’s analgesic effect. The post-treatment times for evaluating pain intensity and pain relief for study endpoints will vary across clinical trials and will be based on the known pharmacodynamic properties of the drug. Assessment of efficacy in the cancer pain population is often more challenging than in the non-malignant pain population. Patients with cancer often suffer with more than one pain syndrome, each syndrome presenting with one or more pain symptoms. Furthermore, cancer pain syndromes and pain symptoms can differ with respect to their underlying pathophysiological mechanisms. For trials involving the evaluation of drugs with a very specific mechanism of action, the use of one global pain
intensity measure for the primary efficacy endpoint may not be sufficiently sensitive to detect an analgesic response. A global pain intensity or pain relief measure may have limited sensitivity in situations where a particularly bothersome pain symptom responds to the drug in a clinically meaningful way, but the patient’s other pain symptoms have not responded due to differences in their underlying pathophysiological mechanisms. Unfortunately, there is currently no assessment tool specifically designed to simultaneously evaluate changes in the intensity of patients’ distinct pains to an intervention. The Neuropathic Pain Scale (NPS) comes closest to achieving this, but it is intended to be used for simultaneously evaluating multiple symptoms of neuropathic pain only (1). Thus, a patient diary can be developed for a clinical trial to include validated pain intensity and or pain relief scales that allow assessment of treatment-related changes in specific pain symptoms over time.

In further consideration of the multidimensional nature of cancer pain and to increase the sensitivity of a clinical trial in detecting clinically meaningful responses to treatment, additional measures of efficacy should be considered for study endpoints, including measures of physical functioning, emotional functioning, and opioid requirements. It is best, however, that the assessment tools chosen to measure these constructs have been validated in the cancer pain population and have been demonstrated to be sensitive to analgesic drug interventions.

See references 2 through 7 for discussions about efficacy measures for analgesic drug clinical trials. In the example below, multiple primary and secondary endpoints have been chosen to increase the sensitivity of the trial to detect a clinically meaningful response to the investigational drug and to better understand its pharmacodynamic properties. Both Day 5 and Day 8 were chosen for the endpoints in an attempt to maximize the difference in the proportion of responders in the active versus placebo groups. Based on previous studies, it is known that drug XXX produces analgesia that persists well beyond the treatment period.

The Primary endpoints include:

a. the proportion of responders in the drug XXX versus placebo groups, where a response is defined as:
   • a ≥ 30% reduction in pain intensity from baseline, on Day 5 and Day 8 post treatment, as measured by the Brief Pain Inventory – Short Form Question #3 (BPI-SF Q#3; numeric rating scale the worst pain in the last 24 h). The reduction in pain intensity must be accompanied by either a decrease or stabilization (<15% increase) in mean opioid analgesic consumption compared with baseline.

b. the proportion of clinical responders in the drug XXX versus placebo groups, where a clinical response is defined as:
   • a ≥ 30% reduction in pain intensity from baseline, on Day 5 and Day 8 post treatment, as measured by the BPI-SF Q#5 (numeric rating scale for average pain in the last 24 hours) if average baseline pain intensity score is ≥4, or
   • a ≥ 30% reduction in pain intensity from baseline, on Day 5 and Day 8 post treatment, as measured by any component-specific pain scale from Patient Diary that has an average baseline pain intensity score of ≥4, or
   • a ≥ 30% reduction in pain intensity from baseline, on Day 5 and Day 8 post treatment, as measured by any component-specific pain measured by the subscales of the Neuropathic Pain Scale (NPS) that has an average baseline pain intensity score of ≥4, or
   • the patient confirms that the global pain (BPI-SF Q#5) or component-specific pain (Patient Diary or NPS) has “very much improved” or “much improved” since the start of the study (Note: these are two categories from a 7-point verbal evaluation scale asking the patient to assess how much their pain has changed since the start of the study, with response categories ranging from 1=very much improved – 7=very much worse).
Safety is assessed through the number of adverse events, the number and nature of abnormal laboratory results, and changes in 12-lead electrocardiogram, blood pressure, heart rate, respiratory rate, and SaO₂.

II.2.C. Secondary endpoints

a. the time of onset of a consistent decrease in pain intensity on the visual analogue scale (VAS) compared to baseline following the first dosing of the treatment phase
b. the post treatment day during which the greatest reduction in pain intensity (BPI-SF Q#3, BPI-SF Q#5, and/or any component-specific pain intensity rating scale) is reported by Responders, or Clinical Responders compared to baseline
c. the interval in days from the first of two consecutive days a patient reports a $\geq 30\%$ reduction in pain intensity (BPI-SF Q#3, BPI-SF Q#5, and/or any component-specific pain intensity rating scale) until the reduction in pain intensity score is $\leq 15\%$ compared to baseline
d. the number of treated breakthrough pain episodes post-treatment
e. the proportion of patients achieving a $\geq 30\%$ of improvement in their general activity (BPI-SF Q#9A) or walking ability (BPI-SF Q#9C)

II.2.D. Exploratory endpoints

a. the proportion of patients with a neuropathic pain component or without a neuropathic pain component who are categorized as responders and non-responders

II.2.E. Study Design

Drug XXX is being evaluated by a multicentre, randomized, double-blind, placebo-controlled, parallel-group trial to evaluate the efficacy and safety of subcutaneously administered drug XXX in patients with inadequately controlled moderate to severe cancer pain. Drug XXX or placebo will be administered b.i.d. to patients over a four day period. The study duration will extend from three to 10 weeks, beginning the first day of screening until the end of patients’ analgesic response.

All concomitant medications taken by the subject from the start of the first treatment period to the completion of the Follow-Up Visit will be documented. The reported medications will be reviewed and evaluated by the Qualified Investigator or designate to determine if they affect the subject’s eligibility to continue to participate in the study.

Screening for Eligibility (Day –28 to –7)

a. Each subject will undergo screening procedures within 7-28 days prior to the baseline period. The following will be required during the screening period to determine eligibility:
b. Informed consent
c. Medical history
d. Review of concomitant medication (including analgesics)
e. Physical and neurological examinations
f. Vital signs (pulse rate, blood pressure, respiratory rate) and body weight
g. 12-lead ECG
h. Laboratory tests (haematology, clinical chemistry; urinalysis)
i. Pregnancy test for women of childbearing potential
j. Review of inclusion and exclusion criteria
k. Characterization of pain and disease:
   • primary cancer site and type
   • identification and characterization of patients’ three most bothersome pains (e.g. location, etiology, pathophysiology, quality, intensity)
l. Patient categorization in part based on the Neuropathic Pain Questionnaire (NPQ) responses (presence versus absence of a neuropathic pain component)
m. BPI-SF
Baseline Period (Day –7 to Day –1):
The duration of the baseline period will be between 5 to 7 days. The following assessments will be
completed once during the baseline period, unless otherwise specified:
   a. Review of medical history from last visit to present date
   b. Review of inclusion and exclusion criteria
   c. Vital signs (pulse rate, blood pressure, respiratory rate) and body weight
   d. Review of concomitant medications
   e. Review of information related to the patients’ three most bothersome pains
   f. BPI-SF (completed daily between 18:00h-21:00h)
   g. Patient Diary (completed daily between 18:00h-21:00h):
      • intensity of three most bothersome pains
      • recording of concurrent medications, including analgesics
   h. Neuropathic Pain Scale (NPS), if applicable

Note: Calculation of average baseline pain intensity will determine final eligibility (the mean “worst” pain
intensity score, calculated from the last five BPI-SF Q #3 scores recorded by the patient during the
baseline period; “moderate” pain will be defined as an average score of 4-5, and “severe” defined as an
average score of 6 or higher).

Treatment Phase (Days 1-4)
Patients will be admitted to the hospital (in-patient) or at the site’s care facility on a daily basis and will be
allowed to leave the facility upon completion of each daily treatment session if judged appropriate by the
investigator. All patients will be randomized on Day 1 to receive a s.c. injection of drug XXX or placebo
twice daily for 4 consecutive days. The first dosing will be given between 8:00-10:00h, and the second
dosing between 14:00-16:00h.

The following will be required prior to each dosing unless otherwise specified:
   a. Review of concomitant medication
   b. Vital signs
   c. VAS for pain intensity (VAS-PI) to help determine acute analgesic response to XXX treatment
      (completed prior to first dosing of Day 1 and Day 4)
   d. 12-lead ECG

Following the first dosing of Day 1 and 4, acute analgesic response will be assessed (VAS-PI) every 15
minutes for the first hour, and then every 30 minutes until the second dosing of the day. A 12-lead ECG
will be completed 1-2 hours after the morning (Days 1-4) and afternoon (Days 2 and 3) dosing. The
following will be required prior to discharge unless otherwise specified:
   a. Vital signs
   b. Brief neurological examination
   c. Review of adverse events/ Adverse event recording
   d. Review of around-the-clock (ATC) analgesics and breakthrough pain medications
   e. BPI-SF (completed daily between 18:00h-21:00h)
   f. Patient Diary (completed daily between 18:00h-21:00h):
      • intensity of three most bothersome pains
      • recording of concurrent medications, including analgesics

Follow-up Visits
All patients will be assessed at the clinic on follow-up Days 5, 8, and 15 and then when pain intensity
returns to baseline levels. Each evening between 18:00h-21:00h during the follow-up period, patients will
complete the BPI-SF, and record in their Patient Diary the intensity of the three most bothersome pains
and concomitant medications. In addition, patients will record their impression of change for each
bothersome pain (from 1 = very much improved to 7 = very much worse) in their Diary on Days 5, 8 and 15. All patients will have the option to enroll in an open label extension protocol on Day 15 or later. The following will be required at each follow-up clinic visit:

a. Physical examination
b. Vital signs
c. Completed BPI-SFs
d. Completed NPSs, if applicable
e. Review of the Patient Diary, including patient’s assessment of change
f. Review of concomitant medication

Whenever Day 15 is the last visit, laboratory evaluations (clinical chemistry, haematology, and urinalysis), 12-lead ECG, and pregnancy test for women of childbearing potential will be completed. Patients who have experienced Severe Adverse Effects (SAEs) or Adverse Effects (AEs) that are at least possibly related to the study medication will be followed-up by telephone on Day 35 to assess their outcome.

If patient’s pain is adequately controlled in the opinion of the investigator and the subject, the following assessments will be completed by telephone or at clinic visits (required at least every two weeks) for up to 6 weeks (Day 43):

a. Patient Diary (weekly post Day 15 until last day of study)
b. BPI-SF (weekly post Day 15 until last day of study)
c. Vital signs (every 2 weeks at a minimum, for a maximum of 4 weeks)
d. NPS, if applicable (at last visit)
e. Laboratory evaluations (clinical chemistry, haematology, and urinalysis at last visit)

II.2.E. Planned sample

Sample size calculation will be based on the primary efficacy endpoint, i.e. comparison of the proportion of responders, based on the worst pain in the last 24 hours from baseline to Day 5 between the XXX and placebo study drug groups. A total of 116 evaluable subjects will be required to detect a difference in proportion of responders between study drug groups of 20% (placebo) versus 50% (XXX) under the following assumptions: Equal numbers of subjects in the two treatment conditions; and 2-sided test, using a significance level of 0.049; and minimum power of 0.90 (90%). A 30% decrease in pain intensity is considered to be a clinically meaningful response (2). The choice of a within-patient 30% reduction in pain intensity is also based on discussion of clinical importance of changes in chronic pain intensity using an 11-point scale (3).

In a population of subjects with painful diabetic neuropathy (8), investigators observed placebo response rates ranged from 10% to 40%, with an average of 26%; it is expected that a placebo response rate for subjects with cancer pain as in the current study will be similar. It has also been shown that the effect of placebo in the treatment of refractory pain in patients with cancer was 18.1% (9). Thus, a response rate of 20% was selected as a reasonable estimate for a placebo effect for this study. If the placebo effect is larger than anticipated (e.g. 30% responder rate in the placebo group), then power is still high enough (87%) to detect the 30% difference in responder rates between the two study groups.

Assuming that 20% of the enrolled patients will discontinue the study or be withdrawn from the study, a total of 146 patients will be required for this parallel-group trial. It is planned to recruit this sample in approximately 25-30 centres across the country with a mean enrolment of 4 patients per centre. Enrolment into the screening phase of the study will be stopped when the anticipated or actual number of subjects has been achieved across all study sites.
II.2.F. Study population
Male or female subjects over 18 years of age with stable but inadequately controlled moderate to severe cancer pain of at least two weeks duration. Patients may experience visceral, somatic and/or neuropathic pain, requiring opioid administration.

II.2.G. Specific inclusion criteria
A subject will be eligible for inclusion in this study only if all of the following criteria apply:

a. Male or female 18 years of age and over;
b. In-patient or out-patient with a diagnosis of cancer;
c. Stable but inadequately controlled pain with current therapy for at least two weeks;
d. Experiencing somatic, visceral and/or neuropathic pain related to cancer;
e. Pain intensity, as assessed by BPI-SF Q#3 meets the definition of “moderate” (score of 4-5) or “severe” (score of 6-10) pain;
f. Life expectancy of > 3 months;

II.2.H. Specific exclusion criteria
For this example, the exclusion criteria have to be adapted to characteristics of the drug, e.g. the pharmacokinetics, pharmacodynamics, potential drug-drug interactions, and adverse effects. A patient will not be eligible for inclusion in this study if any of the following criteria apply:

a. Planned initiation of chemotherapy, radiotherapy, or bisphosphonates within 30 days prior to randomization;
b. Taking lidocaine, mexiletine, or other anaesthetics;
c. History of CO₂ retention, or oxygen saturation (SaO₂) <90% despite O₂ via nasal prongs;
d. Use of scopolamine, acetylcholinesterase-inhibiting drugs, beta-blockers and antiarrhythmic drugs;
e. Second or third degree heart block or prolonged QTc interval (corrected for rate) on screening ECG (confirmed > 470 msec on repeated occasion);
f. Known hypersensitivity to XXX and/or its derivatives;
g. Received an investigational agent within 30 days prior to screening or who is scheduled to receive an investigational drug other than XXX during the course of the study;
h. Females who are lactating or at risk of pregnancy (i.e., sexually active with fertile males and not using an adequate form of birth control);
i. Females with a positive serum pregnancy test at screening or positive urine pregnancy test on admission to study site; or
j. Any other condition that, in the opinion of the investigators, is likely to interfere with the successful collection of the measures required for the study or poses a risk to the patient.

II.2.I. Tools for assessing the endpoints
II.2.1. Tools to assess efficacy
The following table 1 summarizes the assessment tools chosen for this example to measure efficacy variables.

II.2.1.ii. Tools to assess safety
Safety will be evaluated through the assessment of spontaneously reported adverse events, vital signs, physical exam findings and laboratory tests. Other safety data will include 12-lead ECG measurements and brief neurological examinations.

a. Adverse events (AEs)
Adverse events will be recorded daily beginning on Day 1 up to Day 15 for non-serious AEs, and from screening up to 30 days after the last dose of the investigational product.
<table>
<thead>
<tr>
<th>Variable</th>
<th>Assessment Tool Used to Measure Variable</th>
<th>Time of Assessment</th>
</tr>
</thead>
<tbody>
<tr>
<td>Global Pain Intensity in Last 24 Hours</td>
<td>BPI-SF (Q#3, Q#5, Q#6)</td>
<td>Daily through all phases of the trial until Day 15 and then weekly until the end</td>
</tr>
<tr>
<td>Pain Relief in Last 24 Hours</td>
<td>BPI-SF (Q#8)</td>
<td>Daily through all phases of the trial until Day 15 and then weekly until the end</td>
</tr>
<tr>
<td>Component-Specific Pain Intensity</td>
<td>Patient Diary (NRSs; 0=no pain, 10=pain as bad as you can imagine)</td>
<td>Daily through all phases of the trial until Day 15 and then weekly until the end</td>
</tr>
<tr>
<td>Component-Specific Pain Intensity</td>
<td>NPS (subscales)</td>
<td>Baseline (once), Days 5, 8, 15 and last day of trial</td>
</tr>
<tr>
<td>Acute Analgesic Response</td>
<td>VAS-PI</td>
<td>On Day 1 and Day 4, prior and after first dosing (every 15 minutes for the first hour, and then every 30 minutes until the second dosing)</td>
</tr>
<tr>
<td>ATC and Breakthrough Analgesic Use</td>
<td>Patient Diary</td>
<td>Continually through all stages of the trial until the end (morphine equivalents will be calculated)</td>
</tr>
<tr>
<td>Patient’s Impression of Change in Pain</td>
<td>Patient Diary (7-point categorical scale; ‘very much improved’ to ‘very much worse’)</td>
<td>Days 5, 8, 15</td>
</tr>
<tr>
<td>Time to Analgesic Response</td>
<td>BPI-SF (Q#3, Q#5)</td>
<td>Daily through all phases of the trial until Day 15 and then weekly until the end</td>
</tr>
<tr>
<td>Duration of Analgesic Response</td>
<td>BPI-SF (Q#3, Q#5)</td>
<td>Daily through all phases of the trial until Day 15 and then weekly until the end</td>
</tr>
<tr>
<td>Time to Peak Analgesic Response</td>
<td>BPI-SF (Q#3, Q#5)</td>
<td>Daily through all phases of the trial until Day 15 and then weekly until the end</td>
</tr>
<tr>
<td>Walking Ability</td>
<td>BPI-SF (Q#9C)</td>
<td>Daily through all phases of the trial until Day 15 and then weekly until the end</td>
</tr>
<tr>
<td>General Activity</td>
<td>BPI-SF (Q#9A)</td>
<td>Daily through all phases of the trial until Day 15 and then weekly until the end</td>
</tr>
</tbody>
</table>

for severe AEs (SAEs). The frequency of adverse events will be tabulated and summarized according to:
- Type: clinical laboratory abnormalities detected in biological samples, abnormalities detected on physical and neurological examinations, adverse reactions described by the patient.
- Severity: mild, moderate, severe, life threatening
- Association with treatment (causality): probably related, possibly related, not related, unknown.
b. Clinical laboratory evaluations
   Clinical laboratory tests will be performed by a central laboratory. The following evaluations will
   be conducted: clinical chemistry, haematology and coagulation, urinalysis, and pregnancy test.

c. Physical examination, vital signs, body weight, height, brief neurological examination

d. 12-lead ECG
   A 12-lead ECG will be recorded after 10 minutes rest in the supine position at: Screening, Pre-
   dose on Day 1 (randomization; triplicate), 1 hour post dose on Day 1 (triplicate), Pre-dose on Day
   2 (triplicate), 1 hour post dose on Day 2 (triplicate), 1 hour post PM dose on Day 2 (triplicate),
   Pre-AM dose on Day 3 (triplicate), 1 hour post-PM dose on Day 3 (triplicate), Pre-AM dose on
   Day 4 (triplicate), 1 hour post-AM dose on Day 4 (triplicate), and Day 15.

The central reader will measure the electrocardiographic intervals manually on a computer screen using
digital callipers. Each interval will be derived as a mean of three measurements taken from three
consecutive QRST complexes. Mean QT and PR intervals will be used to derive the Bazett (QTcB) and
Fridericia (QTcF) corrected QT intervals. A computer interpretation will be faxed to the site within 30
minutes of transmission. A cardiologist-reviewed ECG report including a full diagnostic interpretation
will be faxed to the site next business day.

II.2.J. Specific criteria for early withdrawal and discontinuation
Subjects will be permitted to leave the study at any time. Subjects can be withdrawn from the study for
any of the following reasons:
   a. Occurrence of a SAE
   b. Administrative reasons (e.g. sponsor decision)
   c. Withdrawal of consent
   d. Major violation of the protocol
   e. Pregnancy
   f. Non-compliance
   g. If it is of the opinion of the Qualified Investigator that it is in the best interest of the patient to
discontinue

Patients discontinuing because of a SAE or because it is in the Qualified Investigator’s opinion that it is in
the best interest of the patient, will be considered to have completed the study. Patients discontinuing
because of withdrawn consent, a major violation of the protocol, pregnancy, or non-compliance will be
non-completers and replaced if they leave prior to drug administration. If a patient is prematurely
discontinued from participation in the study for any reason after drug administration, the investigator must
make every effort to perform the following evaluations: physical examination, 12-lead ECG, vital signs,
clinical laboratory tests including haematology, clinical chemistry, and urinalysis, adverse event
assessment, and pregnancy test (females of childbearing potential). These data should be recorded in the
source documentation and CRF, as they comprise an essential evaluation that should be done prior to
discharging any patient from the study. These subjects will be considered to have completed the study.

In the event that a patient is prematurely discontinued from the study at any time due to an AE or SAE, the
procedures stated in Section XII must be followed. The "End of Study Record" page of the CRF will be
completed for any patient withdrawn from the study.

Patients who drop out of the study due to changing medical status not related to pain, clinically important
changes in non-pain-related medications, or whose status meets one or more exclusion criteria will be
replaced. These patients will not be considered treatment failures.
II.2.K. Data analysis

a. Analysis populations
   All efficacy analyses will be performed using the intent-to-treat principle, i.e. subjects will be
   analyzed based on the study drug group to which they were randomized. All safety analyses will
   be performed for subjects as dosed. In addition to the intent-to-treat analysis approach for efficacy
   (primary), if a substantial proportion of subjects (>10%) fail to complete four days of study
   treatment or there are a substantial number of subjects (>10%) with critical protocol violations
   (e.g. baseline worst pain < 4 for the BPI-SF Q#3), per-protocol analyses will be performed using
   subjects who complete all four days of study treatment with no critical protocol violations.

b. Significance/confidence level
   Differences will be considered statistically significant if the significance level is $\leq 0.049$, 2-tailed.
   Confidence intervals for the absolute difference between the treatment groups in the outcomes
   will be calculated using 95.1% confidence, 2-tailed.

c. Efficacy analysis
   The overall objective is to determine the efficacy of s.c. XXX in reducing the intensity of cancer-
   related inadequately controlled pain compared to placebo.

Primary efficacy analysis
   The primary efficacy analysis will be performed to compare the proportion of patients who are responders
   to XXX with the proportion of patients who are responders to placebo, based on the BPI-SF Q#3.
   Comparison of the proportion who responds in each treatment group will be made using the Mantel-
   Haenszel procedure, stratifying for: baseline pain pathophysiology (includes neuropathic component/does
   not include neuropathic component) and baseline pain intensity as determined by the average of BPI-SF
   Q#3 score (moderate/severe). For the primary efficacy analysis, missing data for Day 5 or Day 8 will
   constitute a non-responder. As a co-primary analysis, a responder analysis will be performed using the
   clinical responder definition of response. The same statistical analysis method will be used for the co-
   primary efficacy analysis as for the primary efficacy analysis. See responder and clinical responder
   definitions on page 4. Exploratory analyses will be performed to compare the mean change from baseline
   in component BPI-SF pain intensity scores and NPS scores using an analysis of covariance (ANCOVA)
   model, with change from baseline in pain as the dependent variable, the study drug group and baseline
   pain pathophysiology as independent variables, and mean baseline pain intensity score as a covariate.
   Least squares means for the study drug groups will be reported.

Secondary efficacy analysis
(i) Determination of whether s.c. XXX treatment improves daily mobility in patients with refractory
   cancer pain.

   The mean change from baseline in BPI-SF Q #9A (general mobility) and BPI-SF Q #9C (walking ability)
   scores will be combined and compared between treatment groups using an analysis of covariance (ANCOVA)
   model, with change from baseline in daily mobility as the dependent variable, the study drug group and baseline
   pain pathophysiology stratum as independent variables, and mean baseline pain intensity score as a covariate.

(ii) Determination of whether s.c. XXX treatment reduces the need for breakthrough medication.
    The overall number of treated breakthrough episodes per patient will be summarized. The number of
    treated breakthrough episodes per patient will be compared between treatment groups using a Poisson
    regression model with the study drug group and pain pathophysiology stratum as independent variables.

(iii) Determination of the onset, duration, and peak of XXX analgesic effect
The onset of analgesia is defined as the first time point showing a consistent decrease in pain intensity, as measured by the VAS-PI, compared with baseline following the first XXX dosing on Day 1 and Day 4.

The duration of analgesic response is defined as the interval in days from the first of two consecutive days a patient reports a \( \geq 30\% \) reduction in pain intensity (BPI-SF Q#3, BPI-SF Q#5, and/or any component-specific pain intensity rating scale) until the reduction in pain intensity score is \( \leq 15\% \) compared to baseline or the patient confirms that the decrease in pain intensity from baseline is no longer meaningful to him or her. The duration of analgesic response may differ depending on the pain component.

Peak analgesic effect is defined as the day during which the greatest reduction in worst pain intensity occurs (BPI-SF Q#3) for responders. Time of peak analgesic effect and duration of analgesic effect will be summarized across treatment groups.

*Exploratory efficacy analysis*

The proportion of responders, using the definition of responder for the primary efficacy analysis will be summarized separately for each pain pathophysiology category (neuropathic, visceral, somatic, and mixed).

d. Safety analysis

*Adverse events*

Adverse event rates will be summarized and compared between the study drug groups for: overall incidence, incidence of related adverse events (according to the causality assessment), incidence of grade 3/4 toxicity per NCI-CTC toxicity criteria, incidence of serious adverse events (including death), and incidence of adverse events leading to discontinuation. Incidence will be counted as treatment-emergent if adverse event onset or worsening occurs after first dose of study drug. For purposes of adverse event analysis, adverse events related to pain will not be summarized.

*Vital signs assessments*

Vital sign data will be summarized at each time point and change from baseline using descriptive statistics (mean, median standard deviation, minimum, and maximum), for systolic blood pressure, diastolic blood pressure, and heart rate. Changes from baseline that exceed the limits indicated in table 2 will be tabulated for each study group.

<table>
<thead>
<tr>
<th>Table 2. Vital sign limits</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Parameters:</strong></td>
</tr>
<tr>
<td>Systolic Blood Pressure:</td>
</tr>
<tr>
<td>Diastolic Blood Pressure:</td>
</tr>
<tr>
<td>Heart Rate:</td>
</tr>
</tbody>
</table>

*ECG assessments*

ECG data will be summarized at each time point (using the mean of 3 observations observed at baseline and 3 observations observed following dosing each day) and reported as change from baseline using descriptive statistics (mean, median standard deviation, minimum, and maximum) for each day for: R-R interval, PR, QRS, QT, and QTc. Regarding QTc, the Bazett and Fredericia corrections for R-R interval will be used. Tolerance limits (Confidence Interval) for the difference between study groups for mean change in QTc at each time point will be estimated. Shift tables will be used to assess baseline versus post-baseline relation of ECG value to reference range for each study drug group, and changes in ECGs from baseline that exceed limits defined in table 3 will be tabulated for each study drug group.
In addition, abnormalities in T wave/U-wave morphology will be summarized for each study drug group.

**Laboratory assessments**

Laboratory data will be assessed using descriptive statistics (mean, median standard deviation, minimum, and maximum) for each of the time points for haematology, blood chemistry, and urinalysis parameters. Shift tables will be used to assess baseline versus post-baseline relation of laboratory value to reference range for each study drug group, and changes to laboratory values that are grade 3 or 4 using the NCI-CTC criteria will be tabulated for each study drug group.

**II.2.L. Adverse events (AEs) and severe adverse events (SAEs)**

The investigator is responsible for the detection and documentation of events meeting the definition of an AE or SAE as provided in this protocol. During the study, when there is a safety evaluation, the investigator or study site personnel will be responsible for XXX AEs and SAEs.

An adverse event (AE) is defined as an unusual and most often undesirable symptom or sign that occurs in human subjects participating in a study. Adverse Events include clinically significant abnormal laboratory values and test results, concomitant illness, accident, medical occurrence or worsening of existing medical condition that emerge during study participation.

All AEs will be recorded on the Adverse Event CRF at each assessment time or when otherwise volunteered by the subject. The investigator will actively solicit this information and assess the AEs from the subject in terms of severity and relationship to study drug. The investigator will treat the subject as medically required until the AE either resolves or becomes medically stable. This treatment may extend beyond the duration of the study. The investigator will record treatment and medications required for treatment on the appropriate CRF(s) and will provide reports of AEs to the sponsor’s clinical monitor on a regular basis during the study conduct.

The investigator will also report to the sponsor all AEs that come to his/her attention after the study termination within 30 days of the last dose of study drug(s).

**III. EXAMPLES OF LANDMARK WELL DESIGNED ANALGESIC TRIALS**

IV. REFERENCES

Patrick du Souich
Professor and chairman of the Department of Pharmacology of the Medical School at the University of Montréal, Québec. Vice-chairman of the Clinical Division of IUPHAR.

Sergio Erill
Director of the Esteve Foundation, a nonprofit institution aimed at fostering communication among scientists. He has been Professor and Chairman of the Department of Pharmacology at two different Spanish universities, and Visiting Professor at the Université de Montreal. He has been president of the Spanish Society for Pharmacology and served in the Selection Committee for Merck Sharp and Dohme International Fellowships in Clinical Pharmacology.

Michael Orme
Emeritus Professor of the University of Liverpool, United Kingdom. He was Professor of Pharmacology and Therapeutics there from 1984 to 2001 during which time he was secretary of the Clinical Section of IUPHAR from 1987 - 1993. He was dean of the faculty of medicine at Liverpool University from 1991 - 1996. He was honorary secretary of the European Association for Clinical Pharmacology and Therapeutics (EACPT) from 1993 - 2003 and is currently chairman of EACPT.
The IUPHAR Compendium of Basic Principles for Pharmacological Research in Humans

Introduction ...............................................................................................................................................4

Chapter 1. Ethical Considerations ...............................................................................................................5
Chapter 2. Good Clinical Practice ...............................................................................................................11
Chapter 3. Assessment of Endpoints: Kinetics and/or Dynamics .............................................................16
Chapter 4. Pharmacogenetics and Pharmacogenomics .............................................................................27
Chapter 5. Paediatric Drug Research .........................................................................................................34
Chapter 6. Phase I (Human Pharmacology) ..............................................................................................40
Chapter 7. Follow-Up of Drugs After Market Entry ..................................................................................50
Chapter 8. Bioavailability and Bioequivalence .......................................................................................55
Chapter 10. Drug Utilization .....................................................................................................................76

Pharmacological Research in Cardiovascular Disorders .......................................................................86

Chapter 11. Hypertensive Vascular Disease .............................................................................................87
Chapter 12. Lipid Lowering Agents ..........................................................................................................106
Chapter 13. Anti-atherosclerotic Drugs .....................................................................................................118
Chapter 14. Heart Failure ........................................................................................................................132
Chapter 15. Arrhythmias ..........................................................................................................................143

Pharmacological Research in Neurologic Disorders .............................................................................164

Chapter 16. Epilepsies and Convulsive Disorders .....................................................................................165
Chapter 17. Headache Disorders .............................................................................................................183
Chapter 18. Alzheimer’s Disease and Other Dementias .........................................................................212
Chapter 19. Parkinson’s Disease and Other Extrapyramidal Disorders ..................................................228
Chapter 20. Multiple Sclerosis and Other Demyelinating Diseases ........................................................248

Pharmacological Research in Mental Disorders ....................................................................................263

Chapter 21. Mood Disorders ...................................................................................................................264
Chapter 22. Anxiety Disorders ...............................................................................................................272
Chapter 23. Schizophrenic Disorders ......................................................................................................278
Chapter 24. Alcoholism and Nicotine Addiction .....................................................................................285

Pharmacological Research in Joint Disorders ........................................................................................293

Chapter 25. Osteoarthritis/arthrosis Short Term Studies ........................................................................294
Chapter 26. Osteoarthritis/arthrosis Long Term Studies ........................................................................308
Chapter 27. Rheumatoid Arthritis ..........................................................................................................318

Pharmacological Research in Other Disorders ....................................................................................337

Chapter 28. Neoplastic Diseases .............................................................................................................338
Chapter 29. Analgesic Drugs for Cancer Pain Management ..................................................................352
Chapter 30. Drugs Used in Osteoporosis .................................................................................................366
Introduction

One of the first recorded human trials was conducted by reverend Edward Stone who found in 50 patients that 1.8 g of powder of willow bark cured their fever, results that were published in 1763 (1). The active compound, salicylic acid, was synthesized only in 1860. Since then, innumerable compounds have been used to cure almost any ailment without evidence of activity. Evidence of drug efficacy was initially required in 1962 with the passing of the Food, Drug and Cosmetic Act by the United States Congress. Currently, in all countries, development and approval of new pharmaceutical entities requires controlled trials proving efficacy. In order to standardize drug registration and approval of drugs, the first International Conference on Harmonization (ICH) was held in 1990. Even if a tremendous progress has been achieved by using ICH guidelines, many aspects of human research remain controversial (2), and even for theoretically rather simple trials, such as those aiming at proving bioequivalence, specifications and study methods differ slightly from one to another in different countries (3).

Should we be concerned with refining the methodology of clinical trials? The answer is yes. Let us consider digitalis. William Withering transformed digitalis from a folk remedy to a modern drug when he transformed a "family receipt for dropsy" that contained more than 20 substances, to a single substance by assuming that foxglove was the active ingredient. Clinical observations enabled Withering to recognize the plant’s narrow margin of safety and the importance of dose: just enough foxglove to cause diuresis, but not enough to cause vomiting or very slow pulse. With these observations, Withering introduced foxglove to the medical profession in 1785 (4). Despite many small trials, it took two centuries to clearly demonstrate the benefits of digoxin in heart failure, and we know now that these benefits include reduction of symptoms, improvement in NYHA class, increased exercise time, modest increased in left ventricular ejection force, enhanced cardiac output, and decreased hospitalizations, and that digoxin does not reduce overall mortality but reduces the rate of hospitalization (5,6).

How to conduct trials to demonstrate drug efficacy? Despite the fact that guidelines for drug development are rather standardized, there is less information about the design of a clinical trial. The objective of this Compendium is to provide the scientific community interested in human research with an easy-to-use reference on how to design a research protocol to assess the effectiveness of a drug in a series of pathological conditions.

The Compendium cannot cover every class of drug and condition, and thus it has primarily focused on cardiovascular and nervous system drugs. The section dealing specifically with the design of clinical trials, chapters 11 to 30, is presented according to a common template to facilitate its consultation. This section is preceded by shorter chapters dealing with general concepts that are applied to the development of almost any drug. The Compendium does not intend to constitute a guideline, but rather an easy source of information on how to design and conduct a clinical trial aiming at demonstrate drug efficacy.

The Editors

Chapter 1. Ethical Considerations

Bozidar Vrhovac, M.D.

Professor Emeritus
Medical School Zagreb
Zagreb
CROATIA
and
Co-chairman
National Bioethics Committee for Medicine of Croatia
and
Corresponding Member
Croatian Academy of Arts and Sciences
I. INTRODUCTORY REMARKS

One of many characteristics of modern society is a pronounced interest in ethical questions. Medicine, especially research on humans, is expectedly, at the top of the list. Why is that so? In spite of the fact that many patients receive therapeutic benefits from participating in clinical trials, benefits that may even be greater than those of standard medical care, randomized clinical trials differ from standard medical treatment in their purpose, characteristics, justification of risks and allocations of interventions according to chance. The research based on various interventions potentially poses risks to the participants that are not always compensated for by medical benefits but that are justified by the potential scientific value of the knowledge which will be got from the trial.

The history of international instruments on ethics is not long. Already before World War II, use of controlled clinical trials was proposed and accepted as the scientific, reliable way of proving efficacy and safety of new therapeutic agents. The atrocious experiments performed by nazi physicians during World War II led, almost immediately (1947) after the war, to the preparation of the Nuremberg Code on ethics of medical research. The Helsinki Declaration followed in 1964 and is now (sixth revision in 2000) taken as the gold standard for research ethics, intending to provide a universal set of principles, which direct the ethical conduct of clinical medical research involving human subjects throughout the world. This is still true in spite of several weaknesses which are at the moment of writing these lines intensively discussed. Other instruments must be mentioned, such as the UN General Assembly Universal Declaration of Human Rights in 1948 and the International Covenant on Civil and Political Rights in 1966. The Belmont Report, elaborated in the US in 1979, is in this country very important for developing new drugs. The Belmont report is, in the US, almost better known than the Helsinki Declaration, and, with its legislative revisions performed later, it is still a very comprehensive instrument.

Recent and very important instruments are the documents issued by the 1990 founded International Conferences of Harmonization (ICH) founded in 1990. Originally the basic aim of ICH was to harmonize the requirements for new drugs in the three biggest drug developers, namely US, European Union and Japan. Later, its guidelines and consensus documents have been accepted by the «rest of the world» most probably because of the importance of the pharmaceutical markets in these countries rather than because of other less materialistic reasons.

The first principle of ICH (taken from WHO GCP in 1995) states: "Clinical trials should be conducted in accordance with the ethical principles that have the origin in the Declaration of Helsinki, and that are consistent with Good Clinical Practice (GCP) and the applicable regulatory requirement(s)".

The booklet prepared by the Council of International Organizations of Medical Sciences (CIOMS) in 1982, in 1993 and in 2002 is today the most informative and comprehensive source of information for ethical research, trying to correct the inconsistencies of the latest revisions of the Helsinki Declaration. These inconsistencies perhaps explain why »studies that breach the provisions of the Helsinki Declaration are still commonly conducted, with the full knowledge of regulatory agencies and institutional review boards».

Of the realistic and justified aspects of the Helsinki Declaration, the most important is the respect for the person’s rights e.g. personhood of subjects, followed by investigators beneficence for subjects participating in the trial and distributive justice in distribution of risk and benefit associated with medical research. The growing importance of persons rights (patients and healthy volunteers) is illustrated by the special attention given to trials in vulnerable and socially unprivileged patients.
Issues of conflict of interest, of transparency and of publishing negative trials are closely linked with ethics as well.

II. THE ETHICS COMMITTEE

The guarantee for the ethical conduct of the study should be a multidisciplinary ethical body called in various countries the Ethics committee, institutional review board, independent ethical committee (EC/IRB). Its size is according to many documents of “at least five” members. This important detail is not mentioned in the Helsinki Declaration. With a small number of members it cannot be expected that an institutional review board will be independent when it decides about resources brought by the sponsor to the institution and its investigators. The number of members must be large enough to ensure that besides the layman, the nurse, the ethicist, and the statistician (who are often named as useful non-scientific members of EC/IRB), at least some members must be experts in the medical and scientific aspects of the clinical trial. Scientific and ethical review cannot be separated. How can someone discuss the ethics of a clinical trial without knowing in detail all facts about the disease in question and its standard treatment? Only medically and scientifically competent members of the ethical committee can safeguard the rights, safety and well-being of the research subjects.

Is it optimal that the same ethics committee, and this is often the case, evaluates research projects and other relevant ethical questions which are constantly present in a health institution, such as artificial prolongation of life of irreversibly sick patients, abortions, unethical behaviour of medical staff, to name only a few.

The EC/IRB should be an independent body, either regional or (for smaller countries) central-national and should discuss research projects only. Such development goes in the described direction and many (even bigger) countries already have central ethics research committees or institutions (US Office for Human Research Protection, U.K. Central Office for Research Ethics Committees, Canada National Council on Ethics in Human Research). In the EU for multicentre trials one member country must give one opinion. This is achieved in various ways one of them being that the central committee delegates the decision to a regional one.

II.1. Informed Consent Document (ICD)

How does the EC/IRB functions and what are the foci of their activity? The most important ethical aspect of the clinical trial is the Informed Consent Document (ICD). The already mentioned International Covenant accepted by the United Nations Assembly in 1966 stresses that « no one shall be subjected without his free consent to medical or scientific experimentation ». A number of documents, meetings and discussions have been written and organized about the optimal format of this important ethical aspect of the clinical trial documents.

It is of the utmost importance that the subject participating in a scientific research project, for instance a clinical trial, understands all details of the planned experiment. To reach this aim the investigator must ensure that the prospective subject has got all the necessary information on the basis of which he reaches at the decision to take part in the trial without having been subjected to coercion, undue influence or inducement, or intimidation. The ICD should contain a statement indicating that the study involves research, should describe the purpose of the research, the expected duration of the subject’s participation, should contain the description of the procedures to be followed (with the indication of which are experimental), of the study treatments, and, if applicable, the nature of random assignment. Moreover the ICD should contain a description of the foreseeable risks or discomforts, of the benefits that may be expected for the subject or for others and disclosure of appropriate alternative procedures or courses of treatment, if any, that might be advantageous to the subject. In addition the ICD will include a statement
describing the extent to which confidentiality and privacy (of records identifying the subject) will be maintained, a statement that participation is voluntary and that the refusal to participate involves no penalty or loss of benefits. It must describe the actions foreseen in the case of injuries (compensation, medical treatment), explanation on whom to contact for additional questions, and, if applicable, any other necessary detail which ensures complete comprehension of above mentioned points by the participant.

Whenever it is not possible to obtain the written ICD, the non-written consent should be documented and witnessed. The problems of obtaining a written ICD for trials in patients with decisional (cognitive) impairment such as patients who are mentally ill, those with Alzheimer’s disease, acutely ill subjects (head trauma, cardiopulmonary arrest, and stroke) or children (no uniform criteria for assent and dissent exist) are considerable. In these cases a legally authorised representative, a proxy or an advanced ICD (in the case of anaesthesia for example) should be used.

In addition, financial details of the trial should be disclosed. Transparency of financial arrangements encourages people to do the right thing. The individual investigator should not stand to benefit personally in financial terms from their involvement in the study.

### II.2. Analysis of Details

Another function of the EC/IRB is to conduct a careful analysis of all details, which could influence the reliability of the trial results. The new sentence in the latest revision of the Helsinki declaration must be mentioned here (articles 19 and 20): “Medical research is only justified if there is a reasonable likelihood that populations in which research is carried out stand to benefit from the results of the research”.

The analysis should begin with the investigator and their team, their potential to recruit subjects without aggressive behaviour, the likelihood of a conflict of interest when trying to serve both the best interests of the patients and the best interests of the research. The analysis should also consider the remuneration and other advantages for both the investigator and their institution. The patient selection, the planned measurements (frequency, justified invasiveness), concomitant and rescue therapy, monitoring of adverse events (especially those which will indicate the need to stop the trial) and comparator therapy must be analyzed.

### II.3. Placebo

The function of the EC/IRB is to analyze the need to use placebo. The use of placebo has been considered by many as non-realistic and unjustified. The most controversial item of the last two revisions of the Helsinki declaration (article 29) states: “The benefits, risks, burdens and effectiveness of a new method should be tested against those of the best current prophylactic, diagnostic and therapeutic methods. This does not exclude the use of placebo, or no treatment, in studies where no proven prophylactic, diagnostic or therapeutic method exists”. Strictly interpreted, this article would rule out the use of placebo controlled clinical trials e.g. a dummy treatment administered to the control group in a controlled clinical trial in order that the specific and non-specific effects of the experimental treatment be distinguished whenever licensed therapeutic method already exists. Active controls (which the investigators are keen for) cannot, in many circumstances, provide reliable evidence of efficacy and safety of the new drug, except by showing the non-inferiority of the new drug. There are many groups of therapeutic agents where placebo controls are justified or even mandatory: analgesics, many psychopharmacologicals, antihypertensives, antianginals, antiarrhythmics and drugs used in primary prevention to name only a few. It is essential that the use of placebo does not pose a risk of serious discomfort, irreversible harm or death or that existing therapy improves survival or decreases serious morbidity.
The sixth revised version of the Helsinki Declaration raised a number of discussions, many more than after the fifth version in which the same proposal was already present, with the result that the World Medical Association prepared a special footnote which states: “The WMA hereby reaffirms its position that extreme care must be taken in making use of placebo-controlled trials and that in general this methodology should only be used in the absence of existing proven therapy. However placebo-controlled trials may be ethically acceptable, even when proven therapy is available, under the following circumstances:
- where for compelling and scientifically sound methodological reasons its use is necessary to determine the efficacy and safety of prophylactic, diagnostic or therapeutic method; or
- where a prophylactic, diagnostic or therapeutic method is being investigated for a minor condition and the patients who receive placebo will not be subject to any additional risk of serious irreversible harm”.

Beside WMA, many meetings in various international organizations like the European Medicines Evaluation Agency (EMEA), the Food and Drug Administration (FDA), the Pharmaceutical Research and Manufacturers of America (PhRMA), the International Federation of Pharmaceutical Manufacturers Associations (IFPMA) and the European Forum for Good Clinical Practice (EFGCP) have discussed the need for a new revision of the Helsinki declaration because of the presence of other conflicting articles. For example, article 30 states that “every patient should be assured of access to the best proven method identified by the study”. This is not acceptable because of the widely known fact that one study cannot identify “the best proven method”.

Therefore, a new version of the Declaration of Helsinki is to be expected. In September 2003 the General Assembly of WMA founded a Working group (its deadline is May 2004) with this goal.

II.4. Payment

Payment in the form of money, gifts, and privileges can only be offered as a recruitment incentive not as a benefit for participation. In clinical trials the prospects of benefit from an experimental treatment and the provision of free ancillary care are viewed as compensation for participation. Healthy volunteers do not need treatment and care. So payment is justified as an incentive for participation.

There are many other ethically sensible areas of clinical research. The examples are, beside those already mentioned, the need to withhold treatment, wash out periods, research involving foetuses and in vitro fertilization, involving pregnant women, children, college students and prisoners. In each of these cases the local EC/IRB has to adapt its decisions according to the local legislation, uses and environment.

In conclusion, clinical trials have numerous ethical aspects. A scientifically and medically well planned clinical trial is ethical and represents the only way for obtaining reliable results which will help in better treatment of a wide circle of patients. Local EC/IRBs have to define what is methodologically essential and ethically appropriate and these aspects are still the subject of intense debate. Ethical committees structured as proposed in this chapter guarantee that ethical principles, accepted to day as appropriate, are observed.

III. SUGGESTED READINGS

11. The European Agency for Evaluation of Medicinal Products: EMEA/CPMP, Position Statement on the USE of Placebo in Clinical Trials with regard to the revised Declaration of Helsinki, EMEA/17424/01.
Chapter 2. Good Clinical Practice

Per Damkier, M.D., Ph.D.

Clinical Pharmacology
Institute of Public Health
University of Southern Denmark
Odense University
Winslewparken 19
5000 Odense C
DENMARK
I. INTRODUCTORY REMARKS

GCP, Good Clinical Practice, is an international set of ethical and quality standards that applies to medicinal trials in humans. Compliance to these standards provides the health authorities as well as the general population with assurance of the integrity of trial subjects and the validity of the data generated. It is important to stress that these standards *de facto* relate to research; in fact a more appropriate title would be Good Clinical Research Practice. These standards have been in effect for many years pertaining to all trials intended to generate data for marketing authorisation procedures whether it be new medicinal products or a line extension for an already marketed product. As a general rule, academic research involving already marketed products and not intended to generate results for marketing authorization purposes has been exempt from these rules. As more general attention and political focus is given to quality assurance and to the autonomy of the individual within health care systems, these standards are getting more attention world-wide; Within the EU, these rules, as of May 1\textsuperscript{st} 2004, pertain to *any* medicinal trials in humans, including those involving already marketed drugs and trials performed without industry engagement. This represents a serious challenge to the academic independent drug related research, as systems to assure GCP compliance must be developed, which in turn requires allocation of appropriate resources.

I.1 History

The first documents describing some quality recommendations for the design of clinical trials in humans can be dated back to USA, where Harry Gold from Cornell University Medical School published two influential papers *Conference on Therapy* in 1946 and again in 1954. The emergence of the thalidomide disaster around 1960 further served to justify the need for formal guidelines and procedures for clinical trials with new medicinal products. Throughout the 60’s and 70’s guidelines were refined and implemented throughout the world. These were, unfortunately, far from being easily comparable due to differences in approaches to clinical trials, mainly between USA, Japan and Europe. It became obvious that some sort of international consensus on this issue was overdue in order to promote mutual recognition of clinical trials and marketing authorization procedures. The result was the birth of the current guidelines that in effect are known as ICH (International Conference on Harmonization)-GCP.

I.2. GCP concepts

In order to give a better understanding of the principal GCP concepts, some key definitions are given below:

*Sponsor*
A person, institution, company or organization which takes responsibility for the initiation, management and/or financing of a clinical trial. Note that the sponsor does not necessarily finance the study. Sponsor and investigator may be identical in which case the term (surprise) *sponsor-investigator* is used.

*Investigator*
A person who is responsible for the trial conduct at the trial site. If multiple investigators are involved at a trial site, a *principal* investigator must me appointed.

*Monitoring*
The act of overseeing a clinical trial to ensure that it is conducted recorded and analysed according to the trial study plan, standard operating procedures, GCP and regulatory requirements.

*SOP, Standard Operating Procedure*
A set of written detailed instructions to achieve uniformity of the performance of a given function.
Audit
An independent examination of all trial related activities and documents in order to determine if the trial was conducted, recorded and analysed according to the trial study plan, standard operating procedures GCP and regulatory requirements. The audit procedure is independent of the monitoring procedure.

II. GCP IN CLINICAL RESEARCH

The ICH-GCP guideline specifies, in rather general terms, how to design, conduct, record and report a clinical trial in accordance with GCP standards. In order to comply with this, it is the responsibility of the sponsor, or sponsor-investigator, to ensure that a set of standard operating procedures is written. An important point is that the guideline explicitly states that the level and intensity of the GCP monitoring process is specified by the sponsor according to the complexity of the study: size, purpose, design, blinding, outcome measures etc. These standard operating procedures form the basis of the central element of GCP-compliance: the monitoring process. The monitoring process must be conducted by individuals possessing documented skills of GCP monitoring, and, obviously, the monitors cannot be directly involved in the study otherwise.

II.1. The monitoring process

The monitoring process is the fundamental aspect of GCP compliance. This systematic process is based on study-specific SOP’s, and serves to document that the study complies with GCP standards. There are three phases of this process some core issues of which are listed below. Please note that this is not an exhaustive listing.

**Before** study initiation the monitor visits each investigation site to verify that
- Sponsor and investigator responsibilities are properly described
- Relevant authorizations are present
- Written informed consent is acceptable
- The investigation site realistically can provide the specified number of subjects within the specified time-frame
- Procedures for handling of trial medicine and laboratory tests are present
- Source documents are specified

**During** the trial, the monitor visits the trial sites to verify that
- The trial is performed and documented as planned
- Informed consent is given from every participant
- In- and exclusion criteria are fulfilled
- Data are correctly recorded and are in accordance with the specified source documents
- Corrections in the case report form (CRF) are properly performed and documented
- Serious adverse event are handled correctly

**After** the trial is completed the monitor verifies that
- Trial medicine is accounted for in detail
- The trial database is properly secured and validated
- Trial documents are filed and stored properly

In effects this means that quite a number of SOP’s must be developed and maintained. A standard operating procedure must exist for all items and procedures below:
- Protocol
- Informed consent
- Investigators brochure
• Case Report Form (CRF)
• Trial medication
• Adverse events
• Protocol amendments
• Monitoring
• Monitor’s report
• Monitor qualification
• Filing
• Audit
• Handling of documents
• Structure and approval of SOP’s

All monitor’s visits must be accompanied by a written report describing what was monitored, the outcome of the monitoring process, including errors and deviations, and initiatives to correct the latter. For the individual researcher this likely represents an insurmountable task and flexible systems to handle GCP in independent academic drug research must be developed.

II.2. A Danish example

In Denmark a public GCP monitoring service has been organised, and a brief overview is given here for inspiration: The first public GCP initiative was taken back in 1995 at the University of Aarhus. Anticipating the implementation of the aforementioned EU directive, it was estimated that about 80 independent academic drug trials in humans were initiated every year in Denmark. A coordinated activity and initiative has resulted in the presence of three GCP-units in Denmark, which are all situated around medical schools and universities. The GCP-units are partly funded by the Government; we are estimating that, having reached steady-state, about 20 full time monitors will be employed, and that a total budget would linger around USD 2 million. The three units work closely together and have agreed upon identical SOP systems, and apply identical principles of services: Smaller trials, typically trials in Ph.D. projects, are monitored free of charge, while larger trials must account for factual GCP-costs, if more than 100 hours of service is required. Despite initial skepticism from researchers this system has so far proven manageable, but the full scale test still waits, at the time of writing, the implementation of the EU directive.

II.3. Challenges and perspectives for independent academic drug research

The challenge is to assure compliance with GCP standards, while not consuming an insurmountable amount of resources, which would cripple the independent academic clinical drug research. It is here that one must recall that the guidelines are general specifications of GCP standards. However, the translation into factual procedures, and the specific way these are implemented leave some room for breathing and interpretation. The approach that the pharmaceutical industry and most CRO organisations have taken is meticulously detailed and very resource consuming. This is understandable, as they cannot afford to have a large pivotal phase III trial be subjected to scrutiny, or even rejection, by health authorities due to a GCP related issue. Hence, the industry developed GCP concept is designed to account for a worst case scenario. So as GCP seems to enter an era were the principle is likely to become applicable to any drug related research in humans, it is really up to all of us, as independent academic clinical researchers, to influence the way that GCP is implemented. There is no denying that implementation of GCP in academic research programs will assimilate some of our scarce funding. However, it is the opinion of this author, that we should welcome many of the principles covered by GCP. Quality in performance and respect for patient’s integrity is imperative for the thrust on which the very existence of any health care system relies, and there are simply no valid arguments to exclude research related activities from these principles. And, if we make our opinion and experience heard it is my belief that it is possible to reach an interpretation of the
GCP guidelines which satisfies this principle and still allows for a fruitful continuation of independent drug research.

III. REFERENCES

The current ICH-GCP guideline can be viewed (among many sites) at the European Medicines Agency’s homepage: www.emea.eu.int/pdfs/human/ich/013595en.pdf.

The FDA has an entire homepage with plenty of easily accessible information: http://www.fda.gov/oc/gcp/default.htm.
Chapter 3. Assessment of Endpoints: Kinetics and/or Dynamics

Rosario Calvo, Ph.D.¹

John C Lukas, M.S., Ph.D. ²

Elena Suarez, Ph.D. ³

Monica Rodriguez, Ph.D. ³

Nerea Leal, Ph.D. ³

¹ Department of Pharmacology
University of the Basque Country
Leioa, Vizcaya 48940
SPAIN
Tel + 34 94 6012761
Fax + 34 94 4800128
Email rcalvo@lg.ehu.es

² Department of Pharmacy and Pharmaceutical Technology
University of Salamanca
Salamanca 37001
SPAIN

³ Department of Pharmacology
University of the Basque Country
Leioa, Vizcaya 48940
SPAIN

Corresponding Author:
Rosario Calvo, Ph.D.
I. INTRODUCTORY REMARKS

A crucial, yet often limiting, factor in the advance of the majority of sciences is the ability to measure and analyze those variables that are truly relevant to a particular field, as for example, in pharmaceutical science, the concentrations in plasma and the effect of a drug. These measurements would be of little use were it not for the development and progress of collateral disciplines such as pharmacokinetics (PK) or pharmacodynamics (PD). The PK examines the relation between the dose administered and the achieved concentration in blood or at the biophase, or “what the body does to the drug”, and the PD deals with the relation between drug amount and effect, or “what the drug does to the body”. In its turn, progress in PK or PD is due mostly to the availability of sensitive and specific analytical methods in order to determine the evolution of the levels of drug and its metabolites in biological fluids and tissues as well as to quantify the drug’s therapeutic effect.

In parallel, the mathematical techniques and methods used in pharmaceutical science to characterize the kinetic processes have also developed considerably, mostly borrowing from other disciplines such as chemistry, physiology or enzymology. The application of these analytical and mathematical techniques to specific studies has enhanced the knowledge of the PK behavior of many drugs, constituting an important advance in adjustment of the dose. But, unpredictability remains a problem since the kinetics usually show large interindividual variation, mainly due to genetic, environmental and pathophysiological factors. For this purpose, state of the art statistical methods generally denominated as “population” PK/PD models have been introduced in the field aiming to evaluate the between (inter) and within (intra) individual differences in the PK/PD in medicated populations of subjects.

The dose is related to the effect through the PK and also through the PD which relates the concentration with the effect. Comprehension of these relations is truly essential for the rational development of therapeutics because the PD determines the target concentration required to produce a specific effect or endpoint, while the PK specifies the appropriate dose regimen to reach that target. This constitutes the basis of what in therapeutics is called “Target Concentration Intervention” (TCI), in contrast to the therapeutic drug monitoring (TDM) methods based solely on a single concentration and Bayes informed PK.

As in the case of PK, methods have also been developed to assess the pharmacological effect in vivo and provide biomathematical model descriptions of the PD. It is already well known that the interindividual differences in PD are numerous and are related to factors such as age, race or disease. Therefore, future efforts must lead to the exploration of covariate models for predicting individual PD parameters. This knowledge together with its PK counterpart could constitute the basis for rational individualized therapy. The combination of PK and PD models permits predicting the temporal evolution of the effect at any dose or regimen.

Also, PK/PD analysis of concentration – time – effect data assists in detecting a series of underlying complications. For example, tolerance development, the formation of active metabolites or the desynchronization between the evolution of the drug concentration and that of the effect. The latter can be due either to drug distributing into an effect compartment or because of an indirect effect mechanism. In all these conditions the plasma concentrations could not be used directly as “targets” (TDM) but integrated PK/PD models would permit estimation of the drug concentration at the effect site as well as the relationship between them.

II. OBJECTIVE

The main objective of therapy is to achieve maximum efficacy avoiding the risk of toxicity. This can be achieved via empirical adjustment of the dose, based either on observations of effects (effect – time
II. Effect – time evolution

The measurement of the pharmacological effect and the adjustment of the dose as a function of the effect appears, in principle, as the most sensible and intuitive approach. However, there are several problems. The observed pharmacological effect, as well as the time it takes to achieve it and its duration, are measurements that change with the dose and the mode of administration, therefore they are variables. This implies that they cannot be used to make predictions in other situations, different from those of the observations at hand.

Additionally, from the knowledge of the effect vs time evolution there can be no direct deduction of parameters, thus having to employ integrated PK/PD analyses as will be discussed below. Parameters are considered those characteristics of the drug that do not change with time, dose or administration route and therefore can be used to adjust or predict adequately any therapeutic regimen. For example, the classical dose - effect relationship provides a useful estimate, that of the dose producing 50% of the maximal effect, ED$_{50}$. Yet, this is not a true parameter since it depends on the time post-dose when the effect was measured.

Dealing with the pharmacological effect is complicated further because of the difficulty to obtain a precise, objective, and continuous measurement of the effect. The use of biomarkers attempts to relieve this problem although we are still far from having the ideal (and validated) biomarker for the majority of drugs.

The use of biomarkers in drug development and clinical practice was revised in the Ninth European Federation of Pharmaceutical Science Conference on Optimizing Drug Development held in Basel 2001 (1). The terminology has been put up to date and the differences between the terms “end point”, “surrogate endpoint” and “biomarker” clarified. Biomarkers are now defined as “physical signs or laboratory measurements that may be detected in association with a pathologic process and that have putative diagnostic and or prognostic utility”. They are, therefore, factors which can be measured objectively and are evaluated as indicators of biological or pathological processes and/or indicators of the response to a therapeutic intervention. In general, biomarkers have a much wider range than surrogate endpoints. We understand as endpoints those variables which can be used to measure how a patient feels after a specific treatment or how a specific body function evolves, “the clinical impact of therapeutic intervention”. The term surrogate endpoint implies that some variables related to the endpoint or the clinical response have been used as biomarkers but are not the final response of the drug (2,3). For example, blood pressure is a biomarker for prevention of hypertension and, at the same time, a surrogate endpoint for the prevention of myocardial infarction and stroke. The measurement of the degree or percentage of prevention would be the endpoint. This example serves also to demonstrate the difficulty in selecting adequate biomarkers since it is still discussed whether the systolic, diastolic or mean blood pressure is the one of interest. The most widely used biomarkers are plasma concentrations of drugs that are used as guides to dosage in clinical practice (e.g. TDM).

Biochemical and molecular biomarkers, such as leukotrienes, angiotensin I and II, or CD4 cell count are of great utility but are hampered by complexity in their mechanism of action which is widely interconnected to other processes. Consequently, no single such biomarker can predict a significant proportion of the observed clinical endpoint. For example, CD4 cell count explains 30% of the survival to HIV and CD4 count plus viral load explains 70%.

Something similar occurs with the gene biomarker products which are also under rapid development. Genes and their function are identified in the genome. Then the proteome is used to identify proteins from
selected genes. Evaluating how mutations cause disease as a result of protein differences, between healthy and diseased subjects, appears to provide candidate gene biomarkers. However, the complexity of the genome or of the pathway from expression to phenotype to macroscopic reality has deflated initial hope.

Adversities aside, the development of adequate biomarkers for a drug apart from better characterization of its PK/PD for analysis or prediction, now appears crucial for the drug development effort. Valid biomarkers help completing the proof-of-concept in the early phases, facilitating decision on continuation with the new drug. Eventually, biomarkers permit reduction of the number of patients in later phases (II or III) and adjustment of the dose in specific populations or in individuals receiving the drug through different routes or dose regimen. Recall however, that even when the ideal biomarker is known, it is necessary to have the PK/PD parameters permitting to make predictions.

II.2. Concentration – time evolution (PK)

The observations of plasma or blood concentration are also non generalizable in their pure form. The maximum concentration reached after a specific dose or the peak or trough after repeated dosing, are measurements useful exclusively for adjusting the dose in situations reproducing the one where they were obtained (body – drug whole). Nevertheless, the measurement of the concentration evolution with time is advantageous, compared to the evolution of the effect, since from this kinetics the PK parameters can be derived that are of great use in adjustment of the dose.

Pharmacokinetics has advanced considerably with the use of mathematical models and computer packages which aid in the analysis and processing of the information generated in clinical practice. The PK also permits the simulation of conditions affecting a particular patient. These models and accessory packages are tools which assist in obtaining parameters but add no scientific surplus value to the information than that input by the experiment, including the user. Modelling will simply reflect the knowledge of the physicochemical characteristics of the drug, the precision of the analytical techniques for drug assays. With more complex models, the qualifications of the user also become more important.

The basic or primary PK parameters are apparent volume of distribution (V) and clearance (CL). V is defined as the relation between dose and initial concentration (V = Dose/ Co). CL is a relation between elimination rate (or distribution rate, for intercompartmental clearances) and concentration in blood or plasma. With passive phenomena (or first order kinetics), which are the most usual in the body, a larger dose implies a larger concentration in blood or plasma and consequently larger elimination rate, thus V and CL are always constant. They are therefore considered parameters and permit prediction of the dose which would be necessary in a patient to produce a specific effective concentration level.

For example, knowing V we could predict the dose needed to reach a specific target (loading dose = target concentration x V). Additionally, CL is a parameter independent of the complexity of the kinetic model (mono, bi or tricompartmental) and is calculated simply as Dose / AUC (assuming complete absorption of the drug). In steady state, after an infusion or multiple dosing, CL is related to the steady state concentration (Css) or average Css (Css = infusion rate / CL), so this parameter can be used to predict the dose regimen (maintenance dose = target Css x CL). Both parameters, CL and V, are primary parameters and are directly related to the physiological processes of the organism. They give us an idea of the relative importance of the space where the drug is distributed and of the organs which eliminate it.

Another PK parameter, commonly employed in dose adjustment, is the elimination rate constant (Kel). It is a mixed parameter that depends on CL and V (Kel = CL / V) and is often recast in the form of the half-life parameter as t1/2 = ln (2) / Kel, now with units of time. This parameter gives an idea of when the steady state is reached and can be used to predict when a Css monitoring sample can be taken.
A facet which should not be neglected when performing a kinetic study or using the concentrations as markers is to know a priori what needs to be finally estimated from the observations. The parameters obtained depend on what has been measured, for example, metabolites, unbound or total drug, enantiomers or racemic mixture.

Another frequently encountered issue is the necessity or not to measure unbound drug (not bound to plasma proteins). It is important to recall that on some occasions the unbound concentration in plasma should be considered since binding in plasma and tissues (the effect site) may not be equal. Nevertheless, if the binding is linear in the range of therapeutic or toxic concentrations, the free and total concentration are simple ratios one of the other and in these cases it is not necessary to measure the free concentration. In contrast, nonlinear plasma protein binding (free drug concentrations increase disproportionately with increasing total drug concentrations) can create havoc in analysis unless free drug concentrations are measured. Perhaps an advantage of parameter estimation as a function of free concentration is that it supplies information about the (intrinsic) behavior of the drug excluding possible differences in the degree of binding. For example, the V of unbound drug, \(V_u\), corrected for the weight, can be extrapolated from animal to man permitting estimation of the first dose in humans. An approximate value of CL for the unbound drug, \(C_Lu\), can be obtained from “in vitro” studies with microsomes in the initial stages of development.

In conclusion, V and CL are fundamental parameters for establishing a dosing regimen, but, depending on the physicochemical characteristics of the drug, the models for the distribution can become complicated and the number of parameters may increase, particularly there may now be more than one volume of distribution. In this case, the volume used to adjust the therapeutic dose, should be the one closer to the effect site (and could also be the steady state volume, \(V_{ss}\)) (see Propofol example below). It is important to remember that the basic concepts reflected in V and CL are always applicable independently of the complexity of the complete model.

**II.3. Programs for data analysis**

Depending on the experiment, the intentions may range from obtaining estimates of PK (PD) model parameters in a single subject or the mean for multiple subjects (a population), or up to the rigorous resolution of the inter and intra subject variability in a population as reflected into statistical distributions of the parameters (Bayesian priors). A subsequent, yet very important task, is usually that of relating the parameters, or their inter subject variability with individual specific covariates, so that a priori prediction of the individual PK (PD) characteristics, and hence the dose, can be improved.

Several software packages have been designed and marketed, mainly in the last 20 years and can be distinguished as single subject analysis programs, or population analysis programs, although both can be used, with the exceptions of some occasions, to perform single or multiple subject analyses. The decision about which approach to use depends, in order of significance, on the density of the observations (rich or sparse data), the scope of the study (e.g. obtaining estimates for a single subject only, or Bayesian priors for a population), and *ad hoc* criteria regarding modelling.

Rich sampling is when there are several drug concentration observations per individual, necessarily more than or equal to the parameters in the model and well distributed in time (e.g. phase I). Sparse sampling is when there are fewer data points per subject than parameters in the PK model to be estimated. Sparse data occur frequently, e.g. in the clinic, due to monitoring restrictions or in phases II, III or IV, due to logistical or ethical considerations. It is also frequent in dose escalation experiments with drug levels below the quantification limit, particularly in small animals, and also in toxicological single point per animal studies.
With rich sampling, individual fitting programs can always be used in single or multiple subject problems, in addition to population specific packages for the latter case (Fig. 1). In sparsely sampled designs, some knowledge about the population at large is always required for estimation of the individual PK model. In single subject sparse PK model estimates, a Bayesian prior from the population has to be introduced at some point in order to inform or support the algorithm on the data gaps. Single subject PK/PD modelling packages are WINNONLIN (Pharsight Corp., Mountain View, CA) and SAAM (Saam Institute Inc., Seattle, WA), this latter permitting the introduction of Bayesian priors from the literature or earlier studies when the data is sparse (e.g. a monitored concentration for each subject).

When many individuals are treated as one within a single subject fit, the analysis performed is known as *naïve* and produces estimates of the mean parameter without any indication of statistical spread. If that information is desired, with rich data and for more than one individual the single subject model fitting run can be simply repeated or iterated for each case. An improvement is a population-like analysis known as standard two stage (STS), because in the first stage all subject specific PK or PD parameters are obtained and in the latter, their centering and dispersion (mean, standard deviation) are simply calculated. With STS, relationships with covariates (e.g. age, weight, sex, and creatinine) can be assessed with standard commercial statistical analysis packages (SPSS, SAS, S-PLUS etc).

*Figure 1.* Schematic of decision tree for use of population methods in data analysis.

For sparse data and multiple subject experiments, the use of *population* or *population modelling* approach is most appropriate. Belonging to the well known statistical problem of *mixed effects*, population algorithms use the information from the remaining population to complete the model of each subject in an iterative process of enriching a prior parameter distribution at each step. Population methods eventually produce estimates of the complete distribution for the PK model parameters (population mean, standard deviation), useful as Bayesian priors, as well as the distribution of the residual error: measures of the inter- and intra-individual variability respectively. Population analysis programs are NONMEM (nonlinear mixed effect modelling, NONMEM Project Group, UCSF, CA) (4) or NPEM (non parametric expectation maximization, Laboratory of Applied Pharmacokinetics, USC, CA) (5) and have been validated and compared (6,7). Recently, a package with a more user friendly interface has been introduced (WINNONMIX, Pharsight Corp., Mountain View, CA), and although not widely used so far, it is lately gaining ground (8,9).
Population fits are usually followed by a maximum a posteriori (MAP) Bayesian estimation step, where the PK (PD) model parameters are obtained for each individual based on the just obtained population priors.

Generally, population analyses are far more complex than single subject approaches, in terms of expertise required and man hours and computer time invested, mainly because there is no single pharmacostatistical model solving the parameter estimation problem. Neither is there a single best fit criterion and the solution process includes visual inspection of residuals or evaluation of confidence intervals. Thus the analyst becomes part of a loop in successive model improvement steps. “Turn of the crank” modelling is impossible with population methods. A simple PK model which may take minutes in a computer and a single subject algorithm to solve may require days or weeks for its population counterpart, even excluding computer time delays due to the size of the sample. The time and knowledge invested is largely extended when multiple occasions of the same subjects and covariate models for the PK parameters are created within the package. The complete population analysis, beginning with the design of the samples, collection, analysis, covariate modelling and possibly simulation, is a highly demanding task.

The PK parameters depend on demographic and physiopathological covariates such as renal insufficiency, diabetes, age, sex and weight and this finally affects the dose. The causality or pathways of such variation are usually not well known and thus, unpredictability remains a problem for most drugs. Covariate model development is a very important effort subsequent to any characterization of the individual kinetics in a population. Some population packages allow introduction of a theoretically unlimited number of covariates in the population fit, thus permitting immediate reduction of the inter-individual variability (NONMEM, WINNONMIX); others allow only a limited number of covariates to be introduced in the fit, thus covariate modelling is performed externally.

Even drugs in use for years, like methadone, in clinical practice, show variability in the response (10). In recent experimental work with methadone, it was observed that sex (11), protein binding (12) and P-glycoprotein (13) modify the PK/PD of methadone and it has been suggested that these covariates could be implicated in the variability observed in the clinic. Another example is propofol. Many studies report on the influence of various covariates on the kinetic parameters of propofol, but without a consensus as to the importance of each one (14). Weight, age and formulation have been associated with the variability. Additionally, plasma protein binding, mainly to lipoproteins, is modified in thyroid, diabetic or critical patients, which could finally impinge on the kinetic parameters (15-17). Models have been developed between these variables and the unbound fraction of propofol “in vitro”, which could aid in the inclusion of lipoprotein levels in population PK analyses (17).

Immunosuppressant medication also shows worrisome variability in the kinetics which is further complicated by the standard oral administration. Much of the variability described in the parameters appears to be associated with the bioavailability which may also vary with post transplantation time (18). In this situation, in addition to estimating the mean parameters it is important to quantify the variability, which is important in adjusting the dose and the regimen. A variation in CL, for example, would immediately reflect in a change in the required maintenance dose or the dosification rate. Population covariate modelling often deals with inter occasion variation within the same subject, in addition to the inter intra individual variabilities.

These concepts are important to the Target Concentration Interval (TCI). In NONMEM, for example, the overall variability for each patient can be summarized in three parts, the between subject variability (BSV) and the inter occasion variability (IOV) for each PK parameter and the within subject variability (WSV) for the concentrations. Then covariate models can be developed to reduce the BSV and IOV. If the dose is to be adjusted between different subjects the BSV and IOV must be treated for the parameter of interest (e.g. CL), but if the adjustment is within the same subject the WSV has to be considered.
The population approach has been used for years, and increasingly, in all phases of drug development (19) and in postmarketing studies. Some of the most recent studies with NONMEM, which have implied an advance in the adjustment of the dose in the clinic, are listed in Table 1.

Table 1. Some of the latest studies where population methods are employed for dose adjustment.

<table>
<thead>
<tr>
<th>Drug</th>
<th>Reference</th>
</tr>
</thead>
</table>

Population methods are intimately tied to simulation, deterministic when random components are absent or stochastic when they are not. The latter is used in drug development for “in silico” dosing regimen and risk assessment, facilitating the reduction of trial patients in the later phases. Most of the above packages can be used to perform simulations in a population setting (20).

The applicability of the models, the presentation of concise final reports, and the relevance of the study are factors which should be considered in order for population analyses to be useful in clinical practice and in learning about the behavior of the drug in its development phases (21).

In conclusion, dose predictions are usually simple applications of the elementary PK principles discussed above and permit estimating a target concentration after administration of a specific dose or a particular regimen. Nevertheless, the target concentration should not be far removed from knowledge of the concentration – effect relationship because, in fact, that is where the target originates. The selection of a target concentration requires exploration of the PD relation not only for the desired effect but the range from undertherapeutic to toxic effects. In clinical practice, there is debate regarding the utility of two approaches: TCI which is based on the above discussed PK/PD based principles and Therapeutic Drug Monitoring (TDM) which is based on adjusting the dose to maintain a range of drug levels, perhaps loosing from sight the concept of target and its immediate relation to an effect.

**II.4. Concentration – effect relationship (PD)**

The concentration – response relationships lead to the estimation of PD parameters, useful in drug development as well as in the clinic. The most common PD parameters are the maximum effect that can be reached ($E_{\text{max}}$) and the $EC_{50}$, or concentration in plasma or blood capable of producing 50% of the maximum effect, known as potency in “in vivo” studies. These parameters are independent of the dose or time and therefore allow prediction of the effect at any concentration independently of how it was generated. These parameters, like with the PK, are obtained from models, although these are typically diagnostic at steady state rather than with explicit time dependence. The classical models are the $E_{\text{max}}$ or hyperbolic model and the sigmoid model which ensue from the classical drug – receptor relations. The $E_{\text{max}}$ model is expressed as,

\[ E = E_{\text{max}} \frac{C}{EC_{50} + C} \]
In this form it is seen that once the parameters are obtained the effect at any concentration (C) can be estimated. Additionally, the parameter EC50, expressed as free (unbound to protein) concentration of drug, offers valuable information since it usually is of the same order of magnitude as the “in vitro” potency of the drug, IC50, obtained from receptor binding studies in early drug development stages. As such, it can be considered as the target for therapy.

In a recent study with lerisetron, a new 5HT3 antagonist in phase III development, the PD was measured using its surrogate effect, the Bezold-Jarish reflex. The observed value for the EC50 “in vivo” was in the range of the affinity of lerisetron in binding studies published earlier, which allowed the assumption that the activity of lerisetron was due to the parent product and not to the presence of possibly active metabolites, as had been suggested for other compounds of the same family (22). This approach has also been used for comparing adenosine and other lipophilic derivatives as well as various benzodiazepines (23,24).

In spite of the importance of knowing the concentration - effect relationship, for many drugs there is little documentation regarding their PD parameters, including those which are commonly monitored such as aminoglycosides, cyclosporin, phenytoin and digoxin. An exception is theophylline for which the Emax value is known (expiratory flow rate) and as well as the EC50. These parameters have been used successfully to estimate the target concentration in the clinic (25).

There is also ample evidence that interindividual differences in PD are sizeable and are associated with variables such as age, race and pathologies. Therefore in the future, effort should focus on resolving the population PD (e.g. with mixed effects approaches) hoping to eventually employ covariates for predicting individual PD characteristics. These studies must be designed in accordance with basic epidemiological principles, i.e. with populations where all the possible covariables (demographic or pathophysiological) can be completed for all subjects. The results from these PD studies together with their PK counterpart would constitute the basis for adequate use of concentrations and effects as therapeutic targets.

II.5. Complex PK/PD situations and applicability of integrated models

The concentration – effect relations can also be used to detect situations where there is an apparent lack of relation between concentrations and effect, when plasma concentrations can not be used directly as biomarkers. Such conditions exist, for example, when there is temporal disequilibrium between plasma and biophase, tolerance, presence of active metabolites, or enantiomers with distinct pharmacological activity, e.g. tramadol (26,27) and methadone (28). But even in these situations, the use of integrated PK/PD models could permit the prediction of the effect, reached after a certain drug dose or through a specific administration route, since the PD diagnoses the concentration necessary for a specific effect and the PK informs us of the dose corresponding to that concentration.

Since the kinetic and dynamic processes are intimately related to the temporal evolution of the pharmacological effect, combined PK/PD models have been developed in place of characterizing separately the concentration vs time and effect vs time relations.

Of all PK/PD models dealing with complicated conditions, the most developed and amply used is the effect compartment model. It resolves the possible disequilibrium between the central distribution compartment and the effect site via an empirical equilibration rate ke0 in a “link” model which allows the estimation of the concentrations at the effect site. In clinical practice, it is used for dose adjustment, particularly in anesthesia with propofol and phentanyl, since they are drugs with complex kinetics of multiple compartments posing the problem, discussed earlier, of choice of appropriate parameter for dose adjustment. For example, with propofol the effective concentration range for hypnosis is 2 - 3 mg/L. If, for
estimation of the therapeutic dose, we were to select the central V in a hypothetical patient of 70 kg weight with V = 10 L, the dose provided would be 30 mg, completely ineffective. If we were to select Vss for the same purpose (Vss = 466 L), the estimate would be 1428 mg, well above the therapeutic dose. The solution lies in performing the same task employing the peak-effect volume (Vpe = 20 L) estimated via the link model. The correct dose would then be 60 mg. This concept is actually programmed into the target controlled infusion pumps used in the operating room for anesthesia with propofol.

Another practical example of the importance of PK/PD integration is evident in a study where two oral formulations of ibuprofen were compared. The PK had been studied in healthy volunteers (typical study of bioequivalence) and the evolution of the effects (fever) in children with hyperthermia. At first, the kinetic study appeared to indicate that the two formulations were different in absorption as reflected in the time needed to assess Cmax (Tmax). Nevertheless, observation of the effect - time evolution in the children did not show any difference. Integrated PK/PD analysis of both populations jointly, with the use of NONMEM, allowed the determination of the causes of this discrepancy. Due to the nature of the indirect response mechanism, via which the fever process proceeds, the differences in the plasma concentration were not reflected in the observed therapeutic response. Eventually, the two formulations were bioequivalent (29).

The integration of PK and PD is key to understanding the use of TCI as an alternative to TDM. This latter approach, widely used at present, often fails precisely because it does not consider the pharmacological effect. The time seems ripe to start paying attention to the concentration effect relation and to think of strategies to individualize the dose with the help of the concentrations but without loosing from sight the synthesis of the PK/PD concepts.

In conclusion, the appropriate combination of biomarker identification and selection, and bioanalytical methods for development and validation and the use of PK/PD models (including population approaches) for fitting data and predicting future clinical endpoints, can provide powerful insights and efficacious guidance for individual patients.

III. REFERENCES

Chapter 4. Pharmacogenetics and Pharmacogenomics

Kim Brøsen, M.D., Ph.D.

Institute of Public Health
Clinical Pharmacology
University of Southern Denmark
Winslowparken 19
DK-5000 Odense
DENMARK
Tel +45 65 50 37 51
Fax +45 65 91 60 89
Email kbrosen@health.sdu.dk
I. INTRODUCTORY REMARKS

It has been recognized for more than 50 years that genetic differences between people contribute to interindividual differences in the response to many commonly used drugs. Pharmacogenetics is the term used for more than 40 years to denote the science about how heritability affects the response to drugs. Pharmacogenomics is an apparently new science about how the systematic identification of all the human genes, their products, interindividual variation, intraindividual variation in expression and function over time may be used both to predict the right treatment in individual patients and to design new drugs. The term pharmacogenomics was coined in connection with the human genome project, but there is no internationally accepted consensus depicting any semantic differences between pharmacogenetics and pharmacogenomics, and in practice the two terms are used interchangeably. However it seems that most would use pharmacogenetics to depict the study of single genes and their effects on interindividual differences in drug metabolising enzymes, and pharmacogenomics to depict the study of not just single genes but the functions and interactions of all genes in the genome in the overall variability of drug response, whether this is caused by pharmacokinetics, pharmacodynamics or both.

The human genome is composed of 3.1 billion nucleotide bases, and the number of genes is about 26,000. Alternative splicing is relatively common and it may add to the complexity of the human proteome. The average gene consists of 3000 bases, but sizes vary greatly, with the largest known human gene being dystrophin at 2.4 million bases. The functions are unknown for over 50% of discovered genes. Another characteristic of the human genome is that by chance two unrelated subjects have approximately 99.9% DNA sequence homology. Still this leaves room for more than 1 million nucleotide bases being different between two randomly selected subjects. Very often the variability in DNA consists in only one nucleotide base change, and if this occurs in more than 100 subjects at a given position in the DNA, it is referred to as a single nucleotide polymorphism, abbreviated SNP and pronounced “snip”. SNPs make up about 90% of all human genetic variation and it occurs at every 100 to 300 bases. Two of every three SNPs involve the replacement of cytosine (C) with thymine (T), and SNPs can occur in both coding (gene) and noncoding regions of the genome. Many SNPs have no effect on cell, organ or whole body function, but some could predispose people to disease or influence their response to a drug.

In the classical pharmacogenetics a genetic polymorphism was a monogenic or Mendelian inherited pharmacological trait existing in at least two phenotypes (and presumably in at least two genotypes) the rarest of which exist in at least 1% of the population. The difference between the classical and the modern definition of a genetic polymorphism is among others, that a phenotype caused by homozygosity of an SNP with a frequency of 1% has a frequency in the population of one in 10,000.

Most of the classical pharmacogenetic entities have been discovered initially through an adverse drug event in a single or just a few subjects. This was followed by speculations into the putative pharmacological mechanism and subsequently into the identification of the target protein and eventually the target gene. By sequencing the target gene from subjects with different phenotypes functionally important SNPs could be discovered, and this would lead to a genotype test and eventually the application of genotyping in clinical practice. The modern pharmacogenomic approach is to reverse the sequence and originate from genotypes or haplotypes and subsequently to study if this in any way is related to differences in either efficacy or adverse drug events in individual subjects in regular clinical studies. To summarize, classical pharmacogenetics has searched the gene(s) for an abnormal drug response of proven clinical value whereas pharmacogenomics presently and in the future will search the bearing if any, on drug response of known genes, SNPs or haplotypes.

During the past 50 years pharmacogenetics has focussed on drug metabolizing enzymes, and there are a number of reasons why this is so. Firstly, drug metabolizing enzymes may vary more in function and expression than most other pharmacokinetic or pharmacodynamic targets, the reason being, that drug
metabolising enzymes as a general rule are not per se related to health or survival. Secondly, provided that the drug is administered intravenously or absorbed completely following oral ingestion which is true for most but not all drugs) the area under the plasma concentration versus time curve (AUC) or the mean steady-state concentration of the drug is a very specific marker of drug elimination, i.e. a low concentration means rapid elimination and a high plasma concentration means slow elimination. If in turn the elimination of drug depends on one enzyme, then the drug level is a very specific in vivo marker of the enzyme in question. Finally, the experimental setup, the drug assay technology and enzyme kinetics and pharmacokinetics applied in both in vitro and in vivo drug metabolism research is very much the same irrespective of what drug metabolizing enzyme or what drug is under study, and that creates a big flexibility in research.

While modern pharmacogenomics still is a science in the making, the pharmacogenetics of drug metabolising enzymes and particular the cytochrome P450 (CYP) enzymes has been in the focus for almost 30 years. This remainder of this chapter deals with in vivo methodologies and strategies in CYP pharmacogenetics.

II. CYP PHARMACOGENETICS

II.1. Cytochrome P450

Most drugs are lipophilic compounds that are mainly eliminated by oxidation catalyzed by the cytochrome P450 (CYP) enzyme system in the liver. The total number CYP genes in all species is 270, but the human CYP superfamily consists of 57 CYP genes and 33 pseudogenes organized in 18 families and 42 subfamilies. The CYP play a key role in the metabolism, and the endogenous substrates for CYP include fatty acids, eicosanoids, sterols and steroids, bile acids, vitamin D, retinoids and uroporphyrins. The CYP also represent the most important way for detoxification of many foreign chemical including drugs. The drug metabolizing CYP belong to families 1, 2 and 3.

II.2. CYP2D6

The CYP2D6 is the source of the sparteine/debrisoquine oxidation polymorphism, and in 7-9 % of Caucasian populations referred to as poor metabolizers the enzyme is not expressed due to mutations in the CYP2D6 gene on the long arm of chromosome 22. The frequency of poor metabolizers in Blacks and Orientals is only about 2-3%. CYP2D6 displays a marked allelic heterogeneity, and approximately 80 known variants, mainly in the form of single nucleotide polymorphisms in the gene, have been reported. However, the CYP2D6*3, *4 and *5 together predict about 90% of poor metabolizers. The CYP2D6 oxidizes approximately 60 drugs including all of the tricyclic antidepressants, some antipsychotics, selective serotonin reuptake inhibitors, antiarrhythmics, beta-adrenoceptor blockers and opiates.

Separation of individuals into the extensive and poor metabolizer phenotype can be done by either phenotyping or genotyping, and each approach has advantages and disadvantages. The classical model drugs for studying CYP2D6 are sparteine and debrisoquine, and they still are superior to all other probe drugs in this regard. The reason is that they both are almost exclusively oxidized by CYP2D6 and the amount of metabolite appearing in the urine reflects CYP2D6 very precisely, provided urine collection has been complete. The second important feature is that the fraction of the two drugs not metabolized is excreted unchanged in substantial amounts via the kidneys. Thus the metabolic ratios debrisoquine/4-hydroxydebrisoquine and sparteine/dehydrosparteine in 8-12 urinary samples following oral ingestion of 10 and 100 mg debrisoquine and sparteine, respectively, provide accurate and very specific measures of CYP2D6 irrespective of any urine loss. The distribution of both metabolic ratios is clearly bimodal, and extensive and poor metabolizers thus are clearly separated. Poor metabolizers have debrisoquine and sparteine metabolic ratios above 12.6 and 20 respectively (antimodes), and extensive metabolizers have
metabolic ratios below these values. However both sparteine and debrisoquine are obsolete drugs and hence no longer manufactured. Dextromethorphan is an alternative, and a dextromethorphan/dextrorphan ratio above 0.3 in most population studies has indicated the poor metabolizer phenotype. The problem with most substrates of CYP2D6 apart from sparteine and debrisoquine, in terms of serving as probe drugs, is that in the absence of CYP2D6 their preferred route of elimination is still oxidation via the same route to the same metabolite as that made by CYP2D6 only this being catalyzed by alternatively low affinity CYPs and at a much slower rate. The alternative CYPs also vary in activity, and although the parent compound/metabolite ratio in plasma or urine indeed reflects CYP2D6 very accurately, it does not separate extensive and poor metabolizers completely as do sparteine and debrisoquine metabolic ratios.

In Caucasians approximately 1% carries one or several extra copies of CYP2D6 and these individuals are always ultrarapid metabolizers. However only about 15% of phenotypically rapid metabolizers arbitrarily defined as subjects that have a metabolic ratio of sparteine below 0.15 have the duplicated allele. The molecular mechanism of ultrarapid metabolism in the remainder 85% is not known.

During the last 20 years the pharmaceutical industry has largely banned the development of CYP2D6 substrates because of the difficulties in handling a polymorphically metabolized drug. However there was a time where it was not common or possible to use in vitro methods (see chapter XX) to detect which CYPs catalyzed a particular drug including CYP2D6. There are three different in vivo methods that can be applied in order to find out if a drug is metabolized by CYP2D6. The phenotyped panel approach implies that the drug in question is given to 6-12 healthy extensive metabolizers and a similar number of poor metabolizers for either sparteine or debrisoquine. The drug under investigation is usually administered as a single oral dose, but sometimes it may be necessary to give it repeatedly in order to measure the steady-state concentration. CYP2D6 substrates are characterized by the fact that the AUC or Css is higher, usually 2-5 times higher in the poor metabolizers compared with the extensive metabolizers and that the total drug clearance is similarly lower in the former compared with the latter. Usually, but not always, the elimination half-life is 2-5 times longer in the poor compared with the extensive metabolizers. An alternative approach is to investigate a randomly selected group of typically 20-30 patients receiving the drug in question for treatment and to correlate their steady-state concentrations with the sparteine, debrisoquine or dextromethorphan metabolic ratios. For CYP2D6 the correlation is positive, i.e. the higher the steady-state concentration the higher the metabolic ratio. The third method is based on the use of a selective potent inhibitor and here quinidine has been the most commonly used. Six to twelve healthy extensive metabolizers take the drug either once or repeatedly and AUC or Css is determined before and during the concomitant intake of quinidine 100-200 mg/day. For a typical CYP2D6 substrate either of these pharmacokinetic parameters increases 2 to 5 times during quinidine. Quinidine is preferred over other inhibitors of CYP2D6 because it is the only one that selectively inhibits CYP2D6 and not other CYPs. The method can be refined by including a group of poor metabolizers as a negative control in whom the plasma concentration does not change during quinidine. Of the three methods, the quinidine inhibition method is the least commonly used because of the risk of proarythmias caused by quinidine.

All three methods can be refined to look at individual pathways by detecting partial formation clearances of drugs in relation to the CYP2D6 phenotype. Research not only has identified most of the CYP2D6 substrates but it has also shown that CYP2D6 polymorphism displays marked dose dependent kinetics for most of its substrates and that it is the source of many important drug-drug interactions when two substrates that both are metabolized by CYP2D6 are co-administered.

II.3. CYP2C19

Approximately 20 years ago a genetic polymorphism in drug oxidation different from the sparteine/debrisoquine polymorphism was discovered through a bimodal distribution of the aromatic 4-hydroxylation of the (S)-enantiomer of the antiepileptic drug mephenytoin. In Caucasians 2-3% are poor
metabolizers. The S-mephenytoin oxidation polymorphism displays marked interethnic variability as 15-20 % of Orientals are poor metabolizers. About 10 years ago it was reported that the CYP2C19 is the source of the S-mephenytoin oxidation polymorphism. CYP2C19 also displays a considerable allelic heterogeneity, and so far 9 different single nucleotide polymorphisms have been reported in the poor metabolizers. However, the CYP2C19*2 and the CYP2C19*3 mutations reflecting G→A SNPs in exon 5 and 4 respectively still account for approximately 90 % of the poor metabolizers. The CYP2C19 is an important mediator of the biotranformation of tertiary amine tricyclic antidepressants (N-demethylation of amitriptyline, clomipramine, imipramine and trimipramine), all of the proton pump inhibitors, the bioactivation of proguanil, moclobemide and several other drugs.

Mephenytoin is the classical model drug used for phenotyping, but the drug is no longer in clinical use and hence it is no longer manufactured. However it is still available for pharmacogenetic research. Following the ingestion of a single oral dose of mephenytoin, urine should be collected for up to 8 or 12 hours. Urine is analysed for (R)- and (S)-mephenytoin, and the ratio between the chromatographic peak areas, the S/R ratio, is determined. The stereoselective metabolism of mephenytoin is abolished in the poor metabolizers and hence the S/R ratio is close to unity. In extensive metabolizers, (S)-mephenytoin is rapidly hydroxylated and (R)-mephenytoin is slowly demethylated. Thus the S/R ratio is less than one in extensive metabolizers and usually is less than 0.1. An acid labile metabolite is formed in extensive metabolizers and this is gradually converted to (S)-mephenytoin even if the urine is kept at -20°C. Thus the longer the sample is kept before the assay the higher the S/R ratio. If the urine sample is treated with acid then the S/R ratio becomes much higher in extensive metabolizers but it does not change in the poor metabolizers. In all samples where the initial S/R ratio is determined to be above 0.5 it is advisable to add acid and repeat the assay. In the poor metabolizers the S/R ratio does not change whereas in the extensive metabolizers it increases by a factor 10 or more.

There is certainly a need for a better model drug than mephenytoin, and omeprazole appears to be a candidate. Following a single oral dose of 20 mg of omeprazole a blood sample is drawn after 3 hours. The omeprazole/5-hydroxyomeprazole ratio in plasma has been reported to be above 7 in the poor metabolizers and below 5 in the extensive metabolizers.

The in vivo methods for determining if a drug is metabolized are not different from what has been described above for CYP2D6 except that the inhibitor method is not used. Drugs such as fluvoxamine and moclobemide are potent inhibitors of CYP2C19 but they are not selective for this CYP. Thus an increase in plasma concentration of a drug during concomitant intake of either of these drugs does not prove that the drug in question is metabolized by CYP2C19.

II.4. CYP2C9

CYP2C9 is a major enzyme catalyzing the biotransformation of warfarin, phenytoin, fluvastatin, several NSAIDs, tolbutamide and other oral antidiabetics. The CYP2C9 also is the source of a genetic polymorphism but contrary to the CYP2D6 and CYP2C19 polymorphism this was not discovered through a bimodal distribution of a metabolic ratio for one of the drugs metabolized by the enzyme. Rather it was discovered by sequencing of the gene and detection of several SNPs.

CYP2C9*1 is the wild-type allele, and besides there are two important single nucleotide polymorphisms the CYP2C9*2 associated with a functionally important Arg144Cys substitution and the CYP2C9*3 associated with another important Ile359Leu substitution. In Caucasians the frequencies of the 6 different genotypes are 65-70 %, 15-20 % and 8-10 % for CYP2C9*1/*1, CYP2C9*1/*2 and CYP2C9*1/*3. The poor metabolizer genotypes CYP2C9*2/*2, CYP2C9*3/*3 and CYP2C9*2/*3 each occur in about 1-2%.
Tolbutamide has been proposed to be a candidate for a model drug to probe for CYP2C9, but it is not possible to separate the 6 common genotypes completely. In one clinical study by Scordo et al. (1) the average maintenance dose of warfarin [mg/week (±s.d.)] to achieve the desired International Normalized Ratio (INR) (a measure of anticoagulant effect) was: CYP2C9*1/*1: 39 (±15), CYP2C9*1/*2: 28 (±11), CYP2C9*1/*3: 21 (±7), CYP2C9*2/*2: 21 (±4), CYP2C9*2/*3: 18 (±9) and CYP2C9*3/*3: 9 (±4). Although this study differs in that it does not measure the pharmacokinetics but rather the dose required in different genotypes to achieve a desired pharmacodynamic end point, it is still a paradigm for studying the clinical relevance of pharmacogenetics in drug metabolism. It clearly shows that the CYP2C9 oxidation polymorphism is an important determinant of the warfarin dose, but it also shows that within each geno- (pheno-) type there is a considerable variability caused by other genes, the constitution of the patient and the environment. CYP2C9 genotyping before treatment with warfarin probably has limited, if any, value in practice.

II.5. Pheno- or genotyping for drug metabolizing enzymes

Pheno- or genotyping as an aid for choosing the optimal dose from the start in theory should be performed if the drug is exclusively metabolized by a single CYP, if the drug has a low therapeutic index and if clinical dose titration is not feasible. For CYP2D6 the possible candidates include tricyclic antidepressants, some antipsychotics, and some antiarrhythmics and for CYP2C9 the possible candidates could be warfarin and phenytoin. However pheno- or genotyping for CYP enzymes has never achieved widespread use in clinical practice because, as explained above, the response is not determined alone by a single enzyme or gene.

The advantage of phenotyping compared with genotyping is, that phenotyping gives an up-to-the-minute account of the CYP in question integrating the result of the genetic, constitutional and environmental influences. The disadvantage is that it involves the ingestion of a model drug which is often obsolete as a therapeutic agent and hence no longer manufactured and also that it is necessary to collect urine or to draw a blood sample. Genotyping separates patients into categories, it is not influenced by environmental or constitutional factors and it only has to be performed once in the life time of the patient.

III. FUTURE DIRECTIONS

On the basis of the last 25 years of intensive studying the pharmacogenomics/-genetics of CYP enzymes a number of general statements regarding the role of genetic factors for variation in drug response can probably be formulated. First until proven otherwise the response to any drug is always determined to some extent by genetic factors. However drug response is never determined by a single gene or by a group of genes alone. It is rather determined by several interacting genes and with important influences from the environment and from the constitution of the patient.

Now that the human genome is known in practically all detail there is widespread optimism, that in the near future, it will be possible to tailor the treatment to the individual patient on the basis of the patient’s genotype. This author does not entirely share this optimism. Genotyping as an aid to select the right dose of the right drug in the individual patient would be of theoretical use if the response is mainly determined by a single gene or a limited group of genes, and if all of the environmental and constitutional influences have a more limited influence, and besides are known in detail and can be measured in the individual patient. It is anticipated that less than 10% of drugs in 10 years from now will be prescribed following a pharmacogenetic test. However by extrapolation from the 25 or more years of experience in the CYP field studying pharmacogenetics/-genomics will lead to new important insights and discoveries that will ultimately lead to the development of new and better drugs and to the rational use of drugs that are already on the market. According to this author’s view it is here that the true importance and benefits of pharmacogenetics lies.
IV. REFERENCES

   This chapter was written on the basis of numerous studies that have not been cited in the text. As a textbook that gives a comprehensive overview and is reasonably well updated the following can be recommended:

   There is a website with a constant update and comprehensive overview:

Chapter 5. Paediatric Drug Research

Imti Choonara, M.D.

Professor in Child Health
Academic Division of Child Health (University of Nottingham)
Derbyshire Children’s Hospital
Derby DE22 3DT
UK
Tel  01332 724693
Fax   01332 724697
Email  imti.choonara@nottingham.ac.uk
I. INTRODUCTORY REMARKS

Studies in Europe (1) and Australia (2) have shown that a significant number of children, both in hospital and the community, receive unlicensed and off-label medicines. Unlicensed medicines are where the medicine has been modified from that specified in its product license (3). This may involve crushing a tablet in order to make it into a suspension so that a young child can take the medicine orally. Off-label medicines are those that are used in a different manner than recommended in the product license. This may involve using the medicine in a different age group, for a different indication, at a different dose or by a different route to that recommended (3).

II. PAEDIATRIC DRUG RESEARCH

II.1. Why?

The licensing process was introduced in response to drug toxicity that affected children and adults. In particular, the Medicines Act (UK 1968) and the Kefauver-Harris amendment (USA 1962) were passed following 1) drug toxicity in the newborn infant – e.g. the grey baby syndrome due to chloramphenicol and 2) drug toxicity in the developing fetus – e.g. phocomelia due to thalidomide. It is ironic that legislation introduced following drug toxicity in the newborn infant and the developing fetus has failed to ensure that medicines used in paediatric patients are fully tested in relation to efficacy and toxicity. The use of unlicensed and off label medicines is thought to be associated with a greater risk of drug toxicity (4).

An additional problem associated with the widespread use of off label medicines is the lack of appropriate formulations for young children. Young children cannot swallow tablets and they therefore require liquid formulations. Recent studies have shown that many young children are prescribed tablets or capsules even though they are too young to swallow them (5).

Drug toxicity in children is different to that in adults (6). One therefore needs to study medicines in paediatric patients in order to prevent future cases of drug toxicity. Different mechanisms of drug toxicity in children are illustrated in Table 1.

II.2. Who?

Paediatric drug research needs to involve the pharmaceutical industry working in partnership with paediatric health professionals. The latter group consists of doctors, pharmacists and nurses with paediatric expertise. Ideally paediatric clinical pharmacologists, who have both the expertise of other paediatric health professionals and an understanding of clinical pharmacology should be involved, especially in relation to the design of the clinical trials.

The pharmaceutical industry has previously been reluctant to become involved in drug research in children. The legislative changes that have been introduced in the USA (the FDAMA and the Pediatric Rule) have provided a significant financial incentive to study medicines (7). Discussions are currently taking place in Europe with regards to introducing some financial incentive for the European pharmaceutical industry.

Table 1  Major adverse drug reactions in paediatric patients

<table>
<thead>
<tr>
<th>Year</th>
<th>Drug / Compound</th>
<th>Age group</th>
<th>ADR</th>
<th>Mechanism</th>
</tr>
</thead>
<tbody>
<tr>
<td>1886</td>
<td>Aniline dye</td>
<td>Neonates</td>
<td>Methaemoglobinemia</td>
<td>Percutaneous absorption</td>
</tr>
<tr>
<td>1956</td>
<td>Sulphisoxazole</td>
<td>Neonates</td>
<td>Kernicterus</td>
<td>Protein displacing effect on bilirubin</td>
</tr>
</tbody>
</table>
1959 Chloramphenicol Neonates Grey baby syndrome Impaired metabolism

1979 Sodium valproate Young children (< 3 years) Hepatic failure Abnormal metabolism?

1980 Salicylate Children Reye's syndrome Unknown

1990 Propofol Children Metabolic acidosis Unknown Dose related?

1996 Lamotrigine Children Skin reactions Unknown Associated with comedication with sodium valproate

(Reproduced with permission from Paed Perinatal Drug Ther (6))

II.3. Where?

One cannot perform a paediatric clinical trial in an adult clinical trials unit. It is accepted that sick children need to be treated by paediatric health professionals. Similarly for clinical trials involving children, paediatric health care professionals need to be involved, ideally within a paediatric unit. It is important to recognise that clinical trials and other aspects of paediatric drug research can be performed in district general hospitals (8). For general paediatric conditions, these units are probably preferable to tertiary centres where one is more likely to see a highly selective patient group that is not representative of children throughout the community.

For those clinical trials that involve children as outpatients, it is important that the children are assessed in a child friendly location, i.e. safe with toys and play therapists available. It may also be appropriate to assess the child at home.

II.4. Which Medicines?

The success of any clinical trial is related to the clinical need for the medicine. Investigators are more likely to participate in a study of a medicine which is likely to result in significant clinical benefit to children than one where there is already satisfactory treatment. The clinical need of the medicine will also be taken into account by the ethics committee. Ethics committees are more likely to recognise that a clinical trial is appropriate in children if there is no current treatment available. This does not mean, however, that the clinical trial will automatically be approved as the design of the study may be inappropriate.

The International Conference on Harmonisation, which includes representatives from the European Medicines Evaluation Agency (EMEA), the Food and Drug Administration (FDA) and Japan have issued ICH E11, Clinical Investigation of Medicinal Products in the Paediatric Population (7). This guidance categorises medicinal products and their value in children into three categories. These are shown in Table 2. The aim is to encourage the study of medicines in the first two groups where there is the greatest potential clinical benefit.

Table 2 ICH Classification of medicinal products for children

<table>
<thead>
<tr>
<th>Medicinal products for diseases predominantly or exclusively affecting paediatric patients</th>
</tr>
</thead>
<tbody>
<tr>
<td>Medicinal products intended to treat serious or life threatening diseases occurring in both adults and paediatric patients for which there are currently no or limited therapeutic options</td>
</tr>
<tr>
<td>Medicinal products intended to treat other diseases and conditions</td>
</tr>
</tbody>
</table>
II.5. When?

The timing of studies in children is clearly dependent upon several factors. These include whether one is dealing with a serious or life-threatening disease for which there is currently no or limited treatment available. In this situation early paediatric studies are essential. However, where existing treatment is available then clinical trials in children should only be conducted after initial safety data has been established in adults.

II.6. Which paediatric patients?

The clinical nature of the medicine will determine whether it needs to be studied in neonates, infants or children. It is important that investigators are aware of the ICH Guidance in relation to the classification of the five different age groups in relation to paediatric patients. These are shown in Table 3.

<table>
<thead>
<tr>
<th>Category</th>
<th>Age</th>
</tr>
</thead>
<tbody>
<tr>
<td>Preterm neonates</td>
<td>&lt;36 weeks gestation, 0 – 27 days</td>
</tr>
<tr>
<td>Full-term neonates</td>
<td>0 – 27 days</td>
</tr>
<tr>
<td>Infants and toddlers</td>
<td>28 days – 23 months</td>
</tr>
<tr>
<td>Children</td>
<td>2 – 11 years</td>
</tr>
<tr>
<td>Adolescents</td>
<td>12 – 17 years</td>
</tr>
</tbody>
</table>

II.7. How?

II.7.A. Study design

The study design is crucial in that a poorly designed study will fail to attract investigators, obtain ethical approval and recruit children. Investigators need to ask the following questions:

a. Which paediatric age group is most likely to benefit from the medicine?

b. Should a placebo be included in the trial design? (Placebo is appropriate if there is no existing treatment for the condition. If however, effective therapy is available, then the use of a placebo is neither appropriate nor ethical.)

c. How will the pharmacodynamic effect be studied?

d. Which pharmacokinetic parameters, if any, need to be determined?

e. What is the likelihood of significant drug toxicity?

Regulatory authorities are more likely to raise questions about clinical trials in children than in adults (9). The duration of clinical trials in adult and paediatric patients is similar (in Finland two thirds completed within 12 months) (9). Clinical trials in healthy adult volunteers, however, are significantly shorter (over two thirds completed within 6 months) (9).

II.7.B. Pharmacokinetics

It is important to ensure that the minimum number of samples, involving the smallest amount of blood possible is collected from each patient. Microassays may need to be developed to measure drug concentrations in small volumes of blood. Information regarding the metabolic pathway and pharmacokinetic parameters in adults is essential before commencing pharmacokinetic studies in paediatric patients.

The use of population pharmacokinetics whereby a larger number of children are involved but fewer samples are collected from each patient should be considered (10). It is usually appropriate to carry out pharmacokinetic studies in a subgroup of the children recruited for the clinical trial. It should not be made
a precondition for entry into the clinical trial as this may result in the loss of a significant number of children from the study.

II.7.C. Non-invasive methods
Consideration needs to be given to the use of non-invasive techniques such as the caffeine breath test. The caffeine breath test has been used as a probe for CYP1A2 enzyme activity (11). It involves the use of a stable isotope of caffeine and the collection of breath samples for two hours after administration of the caffeine. It has been used to study drug interactions (induction and inhibition) and also the effect of disease on drug metabolism11.

II.7.D. Pharmacodynamics
It is often difficult to study pharmacodynamic effect in younger patients. For certain conditions, measuring the pharmacodynamic effect is not difficult, e.g. seizures in patients with epilepsy (12), mortality in children with leukaemia. For other conditions, however, it is more difficult to assess pharmacodynamic effect, e.g. bronchodilators in infants under the age of 18 months, assessing pain relief in pre-verbal children and neonates. There are validated pain scales appropriate for use in paediatric patients of different ages (13). It is, therefore, essential that one uses a validated pain scale if one is studying an analgesic drug.

II.7.E. Pharmacovigilance
Drug toxicity in children is different to that in adults (6). This may be due to impaired drug metabolism or altered protein binding, but may also be idiosyncratic. As the child is developing they may be prone to different toxicities to adults. The principles of pharmacovigilance in children are similar to that in adults. Consideration should be given to setting up an Independent Safety Monitoring Board if there is the potential for significant toxicity.

III. CONCLUSIONS
Paediatric drug research is more difficult than similar studies in adults. Paediatric drug research involves patients whereas many adult studies involve volunteers. It is up to paediatric health professionals and the pharmaceutical industry to work together to ensure that we study the right medicines with an appropriate design to ensure that children receive medicines that are fully evaluated scientifically. Such an approach will increase efficacy and hopefully reduce toxicity.

IV. REFERENCES
Chapter 6. Phase I Studies (Human Pharmacology)

Edward M. Sellers, M.D., Ph.D. ¹

Patrick du Souich, M.D., Ph.D. ²

¹ Ventana Clinical Research Corporation
   and
   Departments of Pharmacology, Medicine and Psychiatry
   University of Toronto
   Toronto, Ontario
   CANADA

² Département de pharmacologie
   Faculté de médecine
   Université de Montréal
   Montréal, Québec  H3C 3J7
   CANADA
**I. INTRODUCTORY REMARKS**

**I.1 Objective and overview of Phase I studies**

Phase I studies are designed to determine the metabolic and pharmacological actions of the drug in humans, the side effects associated with increasing doses, and if possible, to gain early evidence on effectiveness.

Phase I studies are usually conducted in healthy volunteer subjects, but may be conducted in patients when information on effectiveness is not possible in healthy volunteers or when adverse drug effects are potentially important. The total number of subjects included varies with the drug, but is generally in the range of 20 to 80. Being the first exposure of an investigational new drug to humans, the subjects are closely monitored.

Phase I studies can also evaluate drug kinetics and metabolism, and the mechanism of action in humans. The Phase I study must provide sufficient information about the drug's pharmacokinetics and pharmacological effects so to allow the design of Phase II studies.

**II. REQUIRED PRE-CLINICAL SAFETY EVALUATION**

Pre-clinical safety evaluation information is required to estimate an initial safe starting dose for the human trials and the identification of parameters for clinical monitoring for potential adverse effects. The pre-clinical safety studies should be sufficient to characterize potential toxic effects under the conditions of the proposed clinical trial. The following aspects of the drug should be evaluated prior to human exposure (1):

**II.1. Safety pharmacology**

Safety pharmacology aims to assess the effects of the drug on vital functions, such as cardiovascular, central nervous and respiratory systems, of the animal.

**II.2. Toxicokinetic and pharmacokinetic studies**

Information on absorption, distribution, metabolism and excretion (ADME) in animals should be made available to compare human and animal metabolic pathways.

**II.3. Single dose toxicity studies**

The single dose (acute) toxicity of a drug has to be evaluated in two mammalian species prior to the first human exposure. A dose escalation study is considered an acceptable alternative to the single dose design.

**II.4. Repeated dose toxicity studies**

A repeated dose toxicity study has to be determined in two species (one non-rodent). The duration of repeated dose toxicity studies is related to the duration, therapeutic indication and scale of the proposed clinical trial. In principle, the duration of the animal toxicity studies should be equal to or exceed the duration of the human clinical trials. A repeated dose toxicity study for a minimum duration of 2-4 weeks would support Phase I (Human Pharmacology) studies up to 2 weeks in duration. Beyond this, 1-, 3-, or 6-month toxicity studies would support Phase I human clinical trials for up to 1, 3, or 6 months, respectively (1).

**II.5. Other studies required prior to human exposure**

Local tolerance should be studied in animals using routes relevant to the proposed clinical administration. Such studies are particularly important for parenteral products.
Prior to first human exposure, *in vitro* genotoxicity studies for the evaluation of mutations and chromosomal damage are generally needed. However, completed carcinogenicity studies are not usually needed in advance of the conduct of clinical trials unless there is cause for concern (1).

Reproduction toxicity studies should be conducted in accordance to the population that is to be exposed to the drug. There are regional differences in the timing of reproduction toxicity studies to support the inclusion of women of childbearing potential in clinical trials. In Japan, assessment of female fertility and embryo-foetal development should be completed prior to the inclusion of women of childbearing potential using birth control in any type of clinical trial. In the European Union, assessment of embryo-foetal development should be completed prior to Phase I trials in women of childbearing potential and female fertility studies prior to Phase III trials. In the United States, women of childbearing potential may be included in early, carefully monitored studies without reproduction toxicity studies provided appropriate contraception and precautions are taken to minimize risk (1).

### III. EXPERIMENTAL DESIGN

Phase I studies usually have non-therapeutic objectives and as such are conducted in healthy volunteer subjects. The studies may be open, baseline controlled or may use randomisation and blinding. The design depends upon the aim of the study. When the objective is to determine the tolerability of the dose range expected to be used in later clinical studies and to assess the nature of adverse reactions, the study will typically include both single and multiple dose administration and a placebo arm. The protocol should specify the toxicity-based discontinuation of the study, as well as the dose adjustment rules in presence of toxicity (2).

In Phase I studies, pharmacokinetic studies aim to assess the drug’s half-life and clearance to anticipate accumulation of parent drug or metabolites. Pharmacokinetics may be determined as a part of efficacy, safety and tolerance studies or via separate studies. Depending upon the drug, pharmacodynamic studies may be conducted in healthy volunteer subjects or in patients with the target disease. Drug blood concentrations can be related to response (PK/PD studies). Preliminary studies of activity or potential therapeutic benefit may be conducted in Phase I as a secondary objective when the study is conducted in patients (2).

### IV. DETERMINATION OF THE MAXIMUM RECOMMENDED STARTING DOSE (MRSD)

The MRSD for the first human clinical trial is derived from the no-observed adverse effect levels (NOAELs) in the most appropriate species, conversion of NOAELs to human equivalent doses (HED), and application of a safety factor. Pre-clinical toxicology studies will generate basically three types of findings that can be used to determine the NOAEL: a) overt toxicity, e.g., clinical signs, macro- and microscopic lesions, b) surrogate markers of toxicity, e.g., serum liver enzyme levels, and c) exaggerated pharmacodynamic effects (3).

The HED shall be estimated with the NOAEL estimated in the most relevant species. Alternatively, in the absence of data on species relevance, the most appropriate species for deriving the MRSD is the most sensitive species, e.g., the species with the lowest HED. Several factors could influence the choice of the most appropriate species including (3):

- a) differences in ADME of the drug between the species;
- b) evidence indicating that a particular model is predictive of human toxicity; and
- c) limited biological cross-species pharmacologic reactivity of the therapeutic.
The NOAEL for medicaments administered systemically to animals is accurately extrapolated to other species and human when doses are normalized to body surface area (mg/m²). The easiest approach is using the following equation (3):

$$HED = \text{animal NOAEL} \times \left( \frac{W_{\text{animal}}}{W_{\text{human}}} \right)^{(1-b)}$$

Where $W$ is the weight in kg and $b$ (equal to 0.67) is a correction factor used to convert mg/kg to mg/m².

For instance, if the NOAEL of an investigational drug estimated in the relevant species (monkey of 3 kg) is 50 mg/kg, the HED will be:

$$HED = 50 \text{ mg/kg} \times (3 \text{ kg}/65 \text{ kg})^{0.33} = 50 \text{ mg/kg} \times 0.3624 = 18 \text{ mg/kg}$$

Once the HED has been determined, a safety factor is applied to provide a margin of safety. This safety factor allows for variability in extrapolating from animal toxicity studies to studies in humans resulting from (3):

- a) uncertainties due to enhanced sensitivity to therapeutic activity in humans versus animals,
- b) difficulties in detecting certain toxicities in animals,
- c) differences in receptor densities or affinities,
- d) unexpected toxicities, and
- e) interspecies differences in ADME.

In practice, the MRSD for the clinical trial is determined by dividing the HED by a default safety factor of 10.

Under selected circumstances, a safety factor greater than 10 should be used (3):

- a) presence of steep dose-response curve for important toxicities in the most relevant species or in multiple species,
- b) severe toxicity or damage to an organ system,
- c) irreversible toxicity in animals,
- d) non-monitorable toxicity, such as histopathologic changes in animals that are not readily monitored by clinical pathology markers,
- e) presence of significant toxicities without prodromal indicators,
- f) non-predictable and unexplained mortality,
- g) variable bioavailability between species, with poor bioavailability in the test species used to derive the HED,
- h) large variability in doses or AUC levels eliciting a toxic effect,
- i) questionable study design or conduct, such as few dose levels, wide dosing intervals, or large differences in responses between animals within dosing groups,
- j) novel therapeutic targets, and
- k) use of animal models with limited utility to evaluate a medicine because of very limited interspecies cross-reactivity or pronounced immunogenicity, or because its effect is elicited by mechanisms that are not known to be conserved between animals and humans.

For selected medicines where an exhaustive and rigorous toxicologic testing has been made, several circumstances allow the use of safety factors of less than 10 (3):
a) the medicament is a member of a well-characterized class, has a similar metabolic profile and bioavailability, presents similar toxicity across all the species tested including humans, and the drug is administered by the same route, schedule, and duration of administration,
b) the toxicity elicited by the drug is easily monitored, is reversible and predictable, and exhibits a moderate to shallow dose-response relationship with toxicities that are consistent across the tested species, and
c) the NOAEL is estimated based on toxicity studies of longer duration than required for the proposed clinical schedule in healthy volunteers.

V. STUDY POPULATION AND SAMPLE SIZE

Phase I studies are usually conducted in healthy volunteer, male and females older than 18 years. Patients may be included when information on effectiveness is desired and it is impossible to assess in healthy volunteers or when the information obtained is difficult to extrapolate to a patient population, e.g. antibacterials, drugs used in affective disorders, antipsychotic drugs, analgesic drugs, antihypertensives, or when drug adverse effects are potentially important, such as in cancer chemotherapy, anti-HIV drugs, etc (2).

The description of the study population should identify the characteristics that are important to understand to interpret and apply the study results. The description of the study population should identify important inclusion and exclusion criteria, demographic characteristics, baseline values of any clinically relevant variables that would be important to understand the treatment effect, and other characteristics of the population that have implications for the extent to which results can be generalised. The selection of the population, e.g. the inclusion and exclusion criteria, is defined according to the population to be included, e.g. healthy volunteer or patient. The total number of subjects included in Phase I studies varies with the drug, but is generally in the range of 20 to 80 (2).

VI. ROUTE OF ADMINISTRATION AND DURATION OF THE STUDIES

The route of administration used of the study medicine during Phase I studies is usually the intended route for the commercialised drug. Ideally, the duration of drug administration should be equal to the intended duration of drug treatment. For drugs intended for chronic use, this is not feasible and the duration of administration shall be determined based upon the kinetics or the dynamics of the drug, e.g. until steady state is reached or until the maximal effect is achieved.

VII. MONITORING

To assess safety, monitoring of the subjects enrolled in the Phase I study is of utmost importance, and its extent will depend upon the aim of the study. Vital signs, blood chemistry and ECG should be monitored. In addition, depending upon the characteristics of the toxicity of the drug, other parameters of organ function shall be monitored. Whenever the Phase I study aims also to assess effectiveness, drug response will be monitored.

Monitoring is critical to decide when dose adjustment should be done or when the study should be discontinued due to unacceptable toxicity (2).

VIII. EXAMPLE OF A PHASE I EXPERIMENTAL PROTOCOL

The investigational drug XXX is a product that in clinical practice is intended to be administered for four consecutive days. Two Phase I studies were designed, the first one, where the volunteers received a single escalating dose and the second, included repeated dosages for four days.
VIII.1. Study Objective(s)

VIII.1.A. Single escalating dose study
Primary: To assess the safety of single doses of XXX.
Secondary: To gather preliminary information about the subjective, performance-altering, and effects of single doses of XXX.

VIII.1.B. Repeated i.m. doses study
Primary: To determine the maximum tolerated sub-chronic (4 days) daily dose of XXX.
Secondary: To determine the minimum number of individual doses into which that maximum tolerated sub-chronic daily dose of XXX must be divided, and to test whether the dose and schedule so determined can also be tolerated during a treatment period of seven days.

VIII.2. Study Design

VIII.2.A. Single escalating i.m. doses of XXX
There will be a total of up to 72 participants divided into 9 groups, with 8 subjects in each group, 6 of who will receive active drug and 2 of whom will receive placebo (the vehicle in which XXX is normally diluted). The active drug assignment within groups will be randomised and double-blind. Groups receiving higher doses will do so only after the previous group has received the next lowest dose (i.e. in an ascending manner). Each subject will receive only one injection. The study will be terminated at any dose which produces clinically significant adverse effects. The study will last 21 days.

VIII.2.B. Repeated i.m. doses of XXX
This is a randomised, double-blinded, placebo-controlled, single-site tolerance study of XXX. Up to 56 normal, healthy males or females, aged 18 to 45 years will be included in the study. Subjects will be divided into seven groups of eight. The duration of treatment is four inpatient days for the first six groups and seven inpatient days for the last group. After each injection of XXX or placebo is administered, the drug’s objective and subjective effects are monitored.

VIII.3. Study Population
Healthy males or females, ages 18 to 50.

VIII.4. Inclusion Criteria
A subject will be eligible only if all of the following criteria apply:

a. Males or females between the ages of 18 and 50.
b. No clinically important abnormal physical findings at the screening examination.
c. Normal or clinically acceptable ECG.
d. Normal blood pressure (systolic: 90-140 mmHg; diastolic: 50-90 mmHg) and heart rate (40-100 bpm).
e. Body Mass Index of 19.0-29.0 (kg/m²).
f. Ability to communicate well with the investigator and to comply with the requirements of the entire study.
g. Willingness to give written informed consent (prior to any study-related procedures being performed) and to be able to adhere to the study restrictions and examination schedule.
**VIII.5. Exclusion Criteria**

A subject will not be eligible if any of the following criteria apply:

- Administration of any investigational drug in the period 0 to 45 days before entry to the study.
- Use of any prescription medication during the period 0 to 30 days or over-the-counter medication during the 0 to 5 days before entry to the study.
- Donation or loss of greater than 400 ml of blood in the period 0 to 12 weeks before entry to the study.
- Serious adverse reaction or hypersensitivity to any drug.
- Inability to communicate or co-operate with the investigator because of a language problem, poor mental development or impaired cerebral function.
- History of drug dependence (except tobacco) or psychiatric illness within the past 2 years.
- Consumption of alcohol within 24 hours prior to dose administration.
- Females who are lactating or at risk of pregnancy (i.e. sexually active with fertile males and not using an adequate form of birth control).
- Females with a positive serum pregnancy test at screening or positive urine pregnancy test on admission to study site.
- Presence of pain incurred by unknown causes.
- History of asthma or other respiratory disease.
- History of neurologic or neuromuscular disease.
- History of hypotension or cardiovascular disease.
- History of bladder or urethral disease.
- Positive urine drug screen for drugs with a high potential for abuse and low persistence in the urine.
- Inability to refrain from smoking during study days.
- Any other condition which, in the opinion of the investigators, is likely to interfere with the successful collection of the measures required for the study.

**VIII.6. Other Study Eligibility Criteria Considerations**

All detailed information regarding warnings, precautions, contraindications, adverse events, and other significant data pertaining to the study drug being used in this study will be taken into account, whether or not explicitly mentioned in this protocol. Such documents may include, but are not limited to, the Investigator's brochure or equivalent document provided by the Sponsor.

**VIII.7. Study Drugs And Dosages**

**VIII.1. Single escalating i.m. doses of XXX**

The initial starting dose was chosen as 1/20 of the chronic non-toxic dose in Rhesus monkeys (NOAEL = 1 µg/kg). Subjects will be admitted to the overnight facilities of YYY CRC the night prior to the study. On the day of the study, drug or placebo will be injected at 8 AM. Safety, physiologic, subjective, performance and effect measures will be collected as outlined for the next 24 hours. At 8 AM the day following the study day, a medical assessment, laboratory tests and ECG will be conducted prior to subject discharge. Subjects will be reassessed 48 hours after discharge for any late onset adverse events. Figure 1 illustrates the experimental design.

**VIII.2. Repeated i.m. doses of XXX**

The duration of treatment is four inpatient days for the first six groups and seven inpatient days for the last group. The total daily dose of XXX will either be increased or decreased and the number of doses into
which that total daily dose is divided will either remain the same or be diminished (resulting in larger individual doses), both based on tolerance to previously administered dose schedules. The first group will receive the maximal single dose used (48 µg/day) of XXX i.m divided into four equal portions administered at 4-hour intervals, for up to 4 days. The experimental design is illustrated in figure 2.

**VIII.8. MISCELLANEOUS**

The protocol contains information about aspects such as overdose with XXX and toxicity management, concurrent medications and non-drug therapies, XXX management, measurements and evaluations, premature discontinuation, data analysis methods, adverse effects and study administration.

**IX. REFERENCES**

Figure I. Nine groups of 8 healthy volunteers (6 receiving XXX and 2 placebo)
Figure 2. Randomized, double-blind, placebo-controlled study of multiple-dose tolerance of i.m. XXX in healthy volunteers.
Chapter 7. Follow-Up of Drugs After Market Entry

Jacques LeLorier, M.D., Ph.D., FRCPC

Centre de recherche - Hôtel-Dieu
Centre hospitalier de l'Université de Montréal
Pavillon Masson
3850 St-Urbain
Montréal, Québec  H2W 1T8
CANADA
I. INTRODUCTORY REMARKS

There is a general agreement with the fact that our operational knowledge on drugs at the time they enter the market is grossly inadequate. It is also generally accepted that, in most cases, delaying entry into the market by requesting additional animal or clinical studies would not answer the remaining questions and would only delay the patients’ access to useful and sometimes life-saving drugs (1). The solution is therefore to continue studying drugs in a formal way for an indeterminate period of time following their entry into the market (2). Those who consider indeterminate too long a period of time might wish to recall the cisapride experience.

Although few active participants and observers of the medication scene would disagree with the above statements, there is considerable confusion and indecision as to how to proceed in a practical and economical way in order to answer the numerous questions that still remain at the time of entry into the market.

It is important to realise at this point that the question is not necessarily global or universal, and that it has important connotations in regard to specific countries. This is due to the fact that some countries are traditionally allowing drugs into the market sooner than others (3). This analysis will therefore be done from a Canadian perspective, which takes into account the fact that most drugs have been marketed in the United States and/or the European Union from 6 to 12 months before being allowed access to the Canadian market (4-6). Of more than a theoretical interest is the question of whether Canadian patients benefit or suffer from this time lag.

II. WHAT DO WE KNOW OF THE DRUG ON THE DAY IT ENTERS THE CANADIAN MARKET?

Essentially, we have two sources of information. The first being the results of the Phase III trials and the second, the "real life" experience in the countries where the drug was marketed for a significant period of time before Canada. The Phase III trials are designed to demonstrate efficacy and to reassure regulators that the drug doesn't produce clinically severe and relatively frequent adverse events after short periods of exposure. They are done in very artificial settings on very carefully selected patients by very experienced teams of nurses and doctors. As a result of these conditions, they have very high internal validity but very low generalisability; their main drawbacks being the relatively small number of patients (3,000 to 5,000) and their short duration of treatment with the drug in question (6 to 12 months). They are therefore incapable of providing us with critical information on potentially severe side-effects (that occur on less than 1 out of 1,000 patients) that manifest themselves after more than one year of exposure. From a pharmaco-economic point of view, Phase III clinical trials also fail to provide information that is instrumental to the decisions that drug plan managers have to make. Placebo is commonly used as a comparator, whereas what is needed is a comparison with the cheapest or the most frequently prescribed drug for the same indication. In addition, and almost by definition, they provide no information on the most critical question drug plan managers have to answer: How is it going to be used? That is, for which indications other than the officially approved ones, at what doses, in which populations, and most important of all: which drug(s) is it going to replace?

The other source of information usually available in Canada at the time the drug enters the market is the experience gathered from countries where the drug has already been marketed. In terms of safety, several European countries and the United States have reasonably good and spontaneous reporting systems (5,7), which in the recent past have resulted in drugs being withdrawn from these markets before getting Canadian approval. Although it is reasonable to assume that a drug that produces undesirable side-effects in Europeans and/or Americans would also harm Canadian patients, it is much more difficult to extrapolate the drug use data obtained in countries foreign to Canadian settings. The prescription patterns and the availability of competing drugs is usually so different from one country to another that is it extremely risky to predict the Canadian drug use based on data obtained in other countries.
The concept of conditional approval of new drugs, which was discussed by Rawson, West and Appel (2) in a recent paper in the Journal of Clinical Pharmacology, is central to the eventual implementation of the post-marketing study of new drugs. The concept of conditional approval would provide a framework and a structure under which drugs will be studied after their entry into the market. From a Health Canada point of view, these studies would permit a better understanding of the safety of the drug and its long-term efficacy. From the point of view of the drug plan managers these studies would be invaluable in providing information that would allow them to verify whether the drug is used appropriately.

III. TYPES OF PHASE IV STUDIES

Phase IV studies can be divided in four classes: 1. the active pharmaco-vigilance cohort; 2. the prospective efficacy cohort; 3. the simplified clinical trial and; 4. the drug use study.

III.1 The active pharmaco-vigilance cohort

These are prospective patient cohorts who under their most simplified form can be considered as nothing else but patient registers. They allow a large number of patients to be followed for prolonged periods of time. Patients numbering up to 10,000 and treated for five years or more might be necessary to answer some questions, particularly those relating to safety. The object is to collect very specific information with a maximum degree of efficiency. From a practical point of view, this means a maximum of 20 questions with visits being no more frequent than every three months. These studies can be designed to provide additional information on specifically expected problems; for instance on drugs who in Phase III trial would have shown a low incidence of potentially serious problems, like the elevation of liver enzymes, allergic reactions or prolongation of the QT interval, which could be studied in such a cohort and thus provide information on whether these "red flags" will turn out to be predictors of rare but serious adverse events like hepatic necrosis, Stevens Johnson syndrome or sudden death. These prospective pharmaco-vigilance cohorts have to be flexible enough to be able to capture problems that were not suspected from the results of Phase III trials but could be fatal or life-threatening. Ideally, the collection and analysis of the data should be done in "real time" in order to allow a rapid response in cases where serious problems are encountered. These prospective pharmaco-vigilance cohorts have the great advantage of providing solid numerators and denominators, parameters that are notoriously fuzzy when provided by spontaneous case reports. A very practical problem relating to the implementation of an active pharmaco-vigilance cohort is to be able to distinguish them from "seeding" studies, which have contributed to give all Phase IV studies an undeserved bad reputation. The distinction is not that difficult. The implementation of a Phase IV study is legitimate if it answers an important public health question with an appropriate methodology.

III.2. The prospective efficacy cohort

The objective here is to demonstrate long-term efficacy. This is particularly important regarding drugs for which 6 to 12 months trials have demonstrated some degree of efficacy in an indication for which long-term effectiveness is notoriously difficult to obtain. Classical examples would be drugs used to treat obesity or to help in smoking cessation. The logistics can be extremely simple since the purpose would be to define whether the decreases in weight or if the rates of smoking cessation observed at the end of a six month trial are still present after two or three years.

III.3. The simplified clinical trial

This type of trial could be defined as a randomised controlled trial with a maximum degree of freedom as to how the patients are treated once randomisation has been implemented. These are extremely useful when the effectiveness (as opposed to the efficacy) of two drugs needs to be studied. Randomisation
becomes necessary when confounding by indication would create two groups whose clinical characteristics would be so different that the results could not be interpreted without complex, and sometimes not very credible, statistical manipulations. It is important that both patients and physicians are aware of which drug they have been assigned to. It is also critical that patients and their physicians have a maximum degree of leeway as to the frequency of the visits, the collection of laboratory tests and the treatment of adverse reactions from the drugs and/or complications of the disease. One of the main challenges of these studies is to convince the sponsors and the investigators, which have been brought up in the very structured classical Phase III trial, that useful information can be gathered without specific criteria about the frequency of the visits or the way blood pressure should be measured.

III.4. The drug use study

A critical question that drug plan managers would like to have answered at the time they decide on the inclusion of a new drug in their formulary is the following:

a. What is it going to replace?
b. A more specific form of this question would be: is the new drug going to be used as a first line or as a second line treatment?

These two questions are obviously not answered from the data available to the pharmaceutical company at the time the drug enters the market. Very hypothetical and speculative answers can derive from the knowledge of the medical practice, the companies marketing records and the way new drugs of the same class were previously used. The drug plan manager can only speculate on how the new drug will eventually be used. It is therefore perfectly reasonable to require drug use studies as a condition for listing new drugs, including the implicit acceptance by the pharmaceutical company that corrective measures will have to be taken in case the use of the drug happens to deviate excessively from what was predicted to be ideal in cost-efficient therapeutics. In addition, drug use studies can provide very useful information on the doses actually used in medical practice, which on some occasions can be much larger than those used in the clinical trials, and thus invalidate the initial pharmaco-economics studies based on the results of these trials. The appearance of the use of the drug for new, official or unofficial, indications can also dramatically invalidate the predictions of the pharmaco-economics studies based on Phase III data.

When the required information is contained in databases, these should be favoured since studies on databases have the advantage of being relatively cheap and fast. The other advantage of databases is that they permit one to obtain information on a population of doctors and patients who do not know they are being observed, and who therefore operate under their usual behavioural patterns. When the necessary information is not available in databases, it becomes necessary to implement field studies, which have the inconvenience of being much more expensive, but the enormous advantage of being designed specifically to capture all the necessary information. One of the problems with the prospective drug use studies done in the field is that the health professionals and the patients who agree to participate in the study might not be representative of the overall population. Another one is the fact that their knowing they are being observed might cause them to modify the way they practice. These two problems becoming extreme could render useless the interpretation of these studies because they would not represent the "real life" use of the drug under study.

VI. CONCLUSION

In conclusion, given their relative novelty, Phase IV studies constitute an absolute necessity for the protection of the patients and the proper use of the public funds used to reimburse drugs.
V. REFERENCES


Chapter 8. Bioavailability and Bioequivalence

Jake J. Thiessen, Ph.D.

Leslie Dan Faculty of Pharmacy
University of Toronto
Toronto, Ontario M5S 2S2
CANADA
Email jj.thiessen@utoronto.ca
I. INTRODUCTORY REMARKS

Over the last 25 years, Pharmacokinetics has emerged as an integral part of drug development, especially when identifying a drug's biological properties. By pharmacokinetics, one means the application of kinetics to a Pharmakon, the Greek word used to specify drugs and poisons. The term thereby implies the time course and fate of drugs in the body. This general definition broadly embraces absorption, distribution, metabolism (biotransformation) and excretion (ADME). The linking of Pharmacodynamics (response) and pharmacokinetics offers a composite understanding both about how the drug affects the body and how the body affects the drug.

The most comprehensive insight about a drug's inherent pharmacokinetic properties is gained by studying an intravenous dose. This route of administration has the greatest quantitative potential, as it permits a mass balance approach to be applied to distribution, clearance and the body processes associated with excretion and metabolic elimination (e.g. renal, hepatic). The administration of a drug by other routes, notably oral, introduces an uncertainty that reflects the unknown fraction that is actually absorbed. Consequently, such doses alone cannot accurately identify the distribution and clearance processes.

The most important property of any non-intravenous dosage form, intended to treat a systemic condition, is the ability to deliver the active ingredient to the bloodstream in an amount sufficient to cause the desired response. This property of a dosage form has historically been identified as physiologic availability, biologic availability or bioavailability. Bioavailability captures two essential features, namely how fast the drug enters the systemic circulation (rate of absorption) and how much of the nominal strength enters the body (extent of absorption). Given that the therapeutic effect is a function of the drug concentration in a patient's blood, these two properties of non-intravenous dosage forms are, in principle, important in identifying the response to a drug dose. Onset of response is linked to the rate of drug absorption whereas the time-dependent extent of response is linked to the extent of drug absorption. While the bioavailability of each type of non-intravenous product (e.g. oral, inhalation, topical (e.g. patch), rectal, etc.) could be discussed, this chapter will of necessity focus only on orally administered products. They certainly represent the major pharmaceutical class in drug development and patient treatment.

Bioavailability following oral doses may vary because of either patient-related or dosage-form-related factors. Patient factors can include the nature and timing of meals, age, disease, genetic traits and gastrointestinal physiology. The dosage form factors include 1) the chemical form of the drug (e.g. salt vs. acid), 2) its physical properties (e.g. crystal structure, particle size), and 3) an array of formulation (e.g. non-active ingredients) and manufacturing (e.g. tablet hardness) variables.

Not surprisingly, bioavailability is of clinical, academic, and regulatory interest. The latter includes agencies that approve the sale of products in their nation(s), as well as reimbursement agencies. Applications from manufacturers seeking regulatory approval for a new drug (e.g. New Drug Application (NDA)) must furnish exhaustive information about a drug's pharmacokinetics. Typically, such evidence entails studies wherein the drug has been orally administered. While such trials may broadly be viewed as bioavailability studies, many are ostensibly designed to assess the drug's safety and efficacy via strategies of dose escalation and chronic administration. These studies will not be entertained in this chapter. The more pertinent interest in bioavailability relates to questions about absolute extent of absorption (absolute bioavailability), the importance of product formulation changes that are made during a new drug's development process, the comparability of different oral dosage forms (e.g. modified-release versus conventional products), and whether the products can be administered with meals. These facets will receive attention in this chapter.

Manufacturers seeking regulatory approval of competitive (generic) products (e.g. Abbreviated New Drug Application [ANDA]), must provide detailed bioavailability evidence showing head-to-head comparative
performance of their product against the innovator's product. Such trials are fundamentally designed to establish clinical equivalence particularly as it relates to interchangeability or substitutability. Such bioavailability information is particularly germane to this chapter.

**II. COMPARATIVE BIOAVAILABILITY: A UNIVERSAL APPROACH**

Most bioavailability studies, whether for a new or generic product, possess a common theme. A test is conducted to identify the quantitative nature of a specific product comparison. This comparison for a new drug may be, for example, to assess the performance of an oral formulation relative to that of an intravenous dose, or perhaps the performance of a modified-release formulation in comparison to a conventional capsule. For a generic product, it is typically a comparison of a competitive formulation with a reference product. Such commonality surrounding comparative bioavailability studies suggests a universal experimental approach.

All the studies to be described in this chapter basically attempt to establish a drug's concentration versus time profile following product administration in some form of comparative test. As shown in Figure 1, the two primary metrics for such concentration versus time profiles are the area under the curve (AUC) and the maximum observed concentration (Cmax); the former customarily includes the AUC to the last

**Figure 1:** Illustration of the key metrics in a comparative bioavailability trial showing, for example, Test and Reference products. The maximum concentration (Cmax) occurs at the Tmax. The AUC<sub>t</sub> is the total area under the concentration versus time profile to the last sampling time. The area to time infinity (AUC<sub>∞</sub>) is the extrapolated area based on the AUC<sub>t</sub> and the terminal constant (λ<sub>e</sub>).
After obtaining the profiles in a comparative trial, and computing the metrics, conclusions need to be reached regarding the comparison. Statistical methods are applied to test if the metrics are sufficiently similar to be considered equivalent. When the metrics are deemed equivalent, the drug concentration profiles are regarded as fundamentally the same. To achieve this equivalence, the study products' geometric mean ratios (e.g., AUC test/AUC reference), as well as their projected 90% confidence intervals for the population mean ratio, must be located within an 80 to 125% window. For the maximum concentration (Cmax) some regulatory agencies consider it adequate if only the mean ratios are within the interval (See Figure 2).

The preceding universal approach will be recognized as a common thread in the trials to be identified in this chapter. The data requirements for such an approach fundamentally orchestrate the design of the studies, which will be seen to have a rather common or universal format.

III. COMPARATIVE BIOAVAILABILITY STUDIES FOR NEW DRUGS (NDA)

The initial oral formulation for a new drug is frequently used to conduct early human studies of safety and efficacy. Often, early oral bioavailability information about the drug (and this initial formulation) is
obtained by means of studies comparing it with an intravenous dose and/or a solution of the drug. Although such studies will not be described in this chapter, they employ the Universal Approach wherein the comparator is an intravenous dose or perhaps a solution of the drug.

When the oral dosage formulation undergoes changes during the drug development process, the deductive inference concept becomes a helpful tool. It circumvents the need to retest subjects or patients following each formulation change in order to reestablish product safety and efficacy. The fundamental tenet underpinning the logic is similar to that described later for generic product testing. First, it is assumed that the time-dependent drug concentrations in blood from an early formulation are intimately linked with the effects. Second, if a new formulation exhibits the same time-dependent drug concentrations (rate and extent of drug absorption), the new formulation is deemed "bioequivalent" and, by inference, has the same safety and efficacy.

To test reformulated dosage forms, the Universal Approach is employed. The fundamental nature of the study is similar to that described in detail within the "Bioequivalence" section of this chapter.

III.1. Outline of a typical product reformulation bioavailability study:

III.1.A. Objectives
To test the comparative bioavailability of a reformulated and original product and thereby to determine their equivalence.

III.1.B. Primary endpoints
Determine the time-dependent concentrations of the administered drug in the collected blood (or plasma/serum) of each subject following administration of the reformulated and original products.

III.1.C. Secondary endpoints
Determine the time-dependent concentrations of potentially important metabolites (active and contributing to the product's therapeutic response) in the collected blood (or plasma/serum) of each subject following administration of the reformulated and original products.

III.1.D. Exploratory endpoints
Determine the Cmax, AUCt, AUC∞, Tmax, λz and half-life of the primary (and secondary) endpoints following the reformulated and original products, for each subject.

III.1.E. Study design
The study shall be designed in such a way that the effects of formulation can be distinguished from other factors. When two formulations are compared, a randomized two-period, two-sequence crossover study is considered the design of choice. An adequate washout period between periods is needed to avoid drug carryover effects.

All facets of the study are to be tightly controlled. Normally, subjects fast for 10 hours prior to product administration. A dose of the tested product is administered at the start of an experimental day with about 8 ounces (240 mL) of water. Further fluid will be withheld for about 2 hours; standardized meals are permitted beginning at four hours after drug administration. All subsequent meals will be carefully standardized at fixed intervals.

Sequential blood samples (about 12 to 18, including a pre-dose sample) shall be drawn at appropriate, specified, and carefully recorded times (to capture increasing and decreasing concentrations during the absorption, distribution and elimination phases). The collections are to continue for about three terminal drug half-lives in order to capture at least 80% of the total area. At least three to four samples need to be
obtained from the terminal log-linear phase to derive an acceptable estimate of the terminal constant ($\lambda_z$) from linear regression. For long half-life drugs, a truncated AUC (e.g. up to 72 hours) is generally considered adequate.

Blood samples or the harvested plasma/serum shall be analyzed for the administered drug or metabolites by means of a validated analytical method.

III.1.F. Planned sample
The appropriate subject number can be forecast, as described in the "Bioequivalence" section, via the ANOVA error variance associated with the specific metric (e.g. from an earlier study), the expected deviation of the reformulated products' metrics and the bioequivalence criterion (e.g. 90% confidence that the estimated population mean ratio lies between 80 and 125%).

III.1.G. Study population
Healthy volunteers are normally selected, although for some drugs it may, of necessity, be best to conduct the trial in patients. See the "Bioequivalence" section.

III.1.H. Specific inclusion criteria
Healthy males or females will be included in the study population. Preferably, non-smokers will be employed.

III.1.I. Specific exclusion criteria
Women of childbearing potential are to be excluded if there is a potential risk. Other common exclusion criteria are identified in the "Bioequivalence" section.

III.1.J. Tools for assessing primary endpoints
A validated analytical method is needed for both the primary and secondary endpoints.

III.1.K. Specific criteria for early withdrawal and discontinuation
While the number and availability of subjects shall be sufficient to allow all periods of the study to be successfully completed without coercion, subjects shall retain the right to discontinue the trial. Discontinuation reasons may include adverse drug reactions or even personal preferences. All withdrawals must be reported.

III.1.L. Data analysis method:
Consult the Universal Approach (Figure 2) and the "Bioequivalence" section. In summary, ANOVA is to be used to identify the source contributions by factors including subjects, period, formulation and potential interactions. The geometric mean ratio together with the ANOVA residual mean error term are used to identify the statistical basis for the 90% confidence interval for the ratio of the population means (New Formulation/Original Formulation) of the identified metrics (e.g. AUC, Cmax).

III.2. Development of a new formulation (e.g. modified release product)
Delayed-release products typically release the active ingredient at a time later than immediately after administration, thereby sometimes exhibiting an absorption lag time. The first modified-release product requires an NDA. The purpose of the required studies is to determine if the following conditions are met:

a. The drug product meets the controlled release claims made for it;
b. The bioavailability profile rules out the occurrence of what is called "dose dumping", which is the premature release of the drug from the dosage form;
c. The formulation provides consistent performance between individual dosage units;
d. The steady state performance, in comparison to an available conventional product, is equivalent. If, based on accumulated evidence between circulating concentrations of the drug and response, the modified-release product is different, clinical studies will be needed to show the impact of such differences.

The study requirements for modified-release products permit some flexibility, but shall include the following:

a. A single dose crossover comparison of a conventional, immediate release, product and the modified release product (ideally, the study would also include a solution or suspension of the same drug in the same strength);

b. A single dose food-effect study;

c. A steady-state study.

For the first two study requirements above, the Universal Approach is again needed. While summary information about the food effect study is presented here, further details of a food effect study are presented in the "Bioequivalence" section. The steady-state study requirements will not be presented in this chapter because they do not have a similar "comparative" character. The primary requirements for the two comparative studies are:

III.3. Objective A

To test the comparative bioavailability of a modified-release and immediate-release product and thereby to determine their equivalence.

III.3.A. Primary endpoints
Determine the time-dependent concentrations of the administered drug in the collected blood (or plasma/serum) of each subject following administration of the modified-release and immediate-release products.

III.3.B. Exploratory endpoints
Determine the Cmax, AUCt, AUC∞, Tmax, λz and half-life of the primary (and secondary) endpoints following the modified-release and immediate-release products, for each subject. Some agencies will also require the area over the usual dosing interval for the modified-release product.

III.3.C. Study elements
The fasting study shall be designed in such a way that the effects of the formulation can be distinguished from other factors. When two formulations are compared, a randomized two-period, two-sequence crossover study is considered the design of choice. An adequate washout period between periods is needed to avoid drug carryover effects.

Doses are given to subjects following an overnight fast. All remaining aspects are to be controlled as outlined in the "Bioequivalence" section. Data analysis is also conducted as described in that section. However, it should be recognized that differences in Cmax can be anticipated because the fundamental drug release properties for the modified-release and immediate-release products are different. The potential impact of such differences needs to be weighed in the light of concentration versus response evidence.

III.4 Objective B

To test the effect of food upon the bioavailability of a modified-release product.

61
III.4.A. Primary endpoints
Determine the time-dependent concentrations of the administered drug in the collected blood (or plasma/serum) of each subject following administration of the product under fasting and fed conditions.

III.4.B. Exploratory endpoints
Determine the Cmax, AUCt, AUC∞, Tmax, λz and half-life of the primary (and secondary) endpoints following administration of the modified-release product in each period, for each subject.

III.4.C. Study elements
This fasting/fed study shall be designed in such a way that the potential effects of the meal upon the formulation can be distinguished from other factors. When the formulation is tested as required, a randomized two-period, two-sequence, crossover study is considered the design of choice. An adequate washout period between periods is needed to avoid drug carryover effects.

Doses are given to subjects following an overnight fast. In one period, the fast is continued, whereas in the other period a meal is given before dose administration. For meal administration, please refer to the food-effect study described in the "Bioequivalence" section. Typically, both fasting and fed periods become common four hours after dose administration when normal food ingestion cycles are permitted.

All study aspects are controlled as outlined in the "Bioequivalence" section. Data analysis is also conducted as described in that section and under "Universal Approach".

IV. COMPARATIVE BIOAVAILABILITY FOR GENERIC DRUG PRODUCTS (ANDA): BIOEQUIVALENCE STUDIES

The deductive inference concept is also central to bioequivalence testing. The foundation is set, first, through evidence that a specified, approved, reference drug product (e.g. tablet from the innovative manufacturer) has shown acceptable safety and efficacy through an array of clinical trials. Second, a widely held view is embraced that the time-dependent drug concentrations in blood from such a reference product are intimately linked with the therapeutic effects. Third, a principle is adopted, namely that chemically equivalent (same amount of the same active ingredient) and pharmaceutically equivalent products (same dosage form; e.g. conventional tablet), that exhibit the same rate and extent of drug absorption, are bioequivalent. Fourth, bioequivalent products by inference are considered therapeutically equivalent.

When a manufacturer thereby wishes to gain therapeutic equivalence by introducing a competitive generic product into the marketplace, it is not necessary to conduct the full array of trials needed for the first (innovative) product. If equivalence has been demonstrated, according to prescribed study requirements, appropriately determined metrics (Figure 1), and statistical criteria (Figure 2), the generic product by inference is regarded as therapeutically equivalent to the innovative drug product.

The design of and requirements in, bioequivalence studies are fundamentally satisfied through single dose administrations, although there is a lingering interest in multiple dose testing. The focus is on the rate and extent of absorption of the active ingredient, although some jurisdictions (e.g. FDA) continue to show an interest in the primary active metabolite(s). In some cases, notably drugs that exhibit non-linear pharmacokinetics, the dose strength to be tested may be dictated by whether the drug's non-linearity is attributable to the absorption or elimination phase (Health Canada). As a general principle, the studies are designed to test inherent product absorption properties. Thereby, the trials generally specify healthy normal controls that exhibit circumscribed demographics.
Comparative evidence may require not only studies in a fasting condition, but following a specified meal. The latter permit drug formulations to be evaluated under "stressed conditions". If it is shown that competitive products are bioequivalent under both fasting and fed conditions, there is greater confidence that they are therapeutically equivalent when used in patients.

**IV.1 Testing competitive (generic) products under fasting conditions**

The following describes the requirements for most orally administered products, including tablets, capsules and modified-release dosage forms. Nevertheless, it is best to check with each regulatory agency regarding current or special drug- or product-specific requirements.

The bioequivalence study conclusions are commonly extended to all strengths of the products provided the active and inactive ingredients conform to regulatory requirements of proportionality. When these requirements are violated, representative strengths of each formulation type shall be tested.

**IV.1.A. Objectives**
To test the comparative bioavailability of a test and reference product and thereby to determine their equivalence.

**IV.1.B. Primary endpoints**
Determine the time-dependent concentrations of the administered drug in the collected blood (or plasma/serum) of each subject following administration of the test and reference products.

**IV.1.C. Secondary endpoints**
Determine the time-dependent concentrations of potential important metabolites (active and contributing to the product's therapeutic response) in the collected blood (or plasma/serum) of each subject following administration of the test and reference products.

**IV.1.D. Exploratory endpoints:**
Determine the Cmax, AUCt, AUC∞, Tmax, λz and half-life of the primary (and secondary) endpoints following each of the test and reference products, for each subject.

**IV.1.E. Study design**
The study shall be designed in such a way that the effects of formulation can be distinguished from other factors. If two formulations are compared, a randomized two-period, two-sequence crossover study is considered the design of choice. An adequate washout period between periods is needed to avoid drug carryover effects. Replicate studies, although not mandated, offer the advantage of providing a comparison of intra-subject variances for the test and reference products.

All facets of the study are to be tightly controlled. The full characteristics, including lot numbers and expiry dates of the test and reference products, shall be known. Normally, subjects fast for 10 hours prior to product administration. Normally, the highest safe strength/dose of the test or reference product will be administered at the start of an experimental day with about 8 ounces (240 mL) of water. Further fluid will be withheld for about 2 hours; standardized meals are to be permitted beginning at four hours after drug administration. All subsequent meals will be carefully standardized according to a fixed schedule.

For most drugs, subjects shall not be permitted to recline until at least two hours after product ingestion. Physical activity and posture is to be standardized to limit variable effects on gastrointestinal blood flow and motility. Blood samples (about 12 to 18, including a pre-dose sample) shall be drawn at appropriate, specified, and carefully recorded times (to capture increasing and decreasing concentrations during the absorption, distribution and elimination phases). The collections are to continue for about three terminal
drug half-lives in order to capture at least 80% of the total area. At least three to four samples need to be obtained from the terminal log-linear phase to derive an acceptable estimate of the terminal constant ($\lambda_z$) from linear regression. For long half-life drugs, a truncated AUC (e.g. up to 72 hours) is generally considered adequate.

Blood samples or the harvested plasma/serum shall be analyzed for the administered drug or metabolites by means of a validated analytical method.

IV.1.F. Planned sample
While most jurisdictions support a minimum of 12 subjects in a bioequivalence trial, the likelihood of a successful outcome is improved with an increase in the subject number. The appropriate subject number can be forecast via the ANOVA error variance associated with the specific metric (e.g. from published data or a pilot study), the expected deviation of the test product's metric from that of the reference product (e.g. 0.05) and the bioequivalence criterion (e.g. 90% confidence that the estimated population mean ratio lies between 80 and 125%).

IV.1.G. Study population
To minimize variability and focus on the comparison of the two formulations, healthy volunteers are to be selected, although for some drugs it may, of necessity, be best to conduct the trial in patients. Subjects will ordinarily be between 18 and 55 years of age and within the accepted normal range for Body Mass Index. Clinical laboratory tests, notably to assess cardiac, renal and hepatic function, are to be normal based on subject screening. Furthermore, subjects will have undergone an extensive review of medical history and received a comprehensive medical examination.

IV.1.H. Specific inclusion criteria
Healthy males or females will be included in the study population. Preferably, non-smokers will be employed.

IV.1.I. Specific exclusion criteria
Women of childbearing potential are to be excluded if there is a potential risk. Subjects shall not have a history of alcohol or drug abuse. Subjects shall not be receiving drugs for any medical condition. There is to be no known allergy to the administered drug or formulation. As a rule, alcoholic beverages and over-the-counter drugs shall be avoided during the days immediately preceding a trial and for an appropriate interval during the active sample collection period of the trial.

IV.1.J. Tools for assessing primary endpoints
A validated analytical method is needed for both the primary and secondary endpoints.

IV.1.K. Specific criteria for early withdrawal and discontinuation
While the number and availability of subjects shall be sufficient to allow all periods of the study to be successfully completed without coercion, subjects shall retain the right to discontinue the trial. Discontinuation reasons may include adverse drug reactions or even personal preferences. All withdrawals must be reported.

IV.1.L. Data analysis method
All study information, including exploratory endpoints shall be presented for each subject following the test and reference products. ANOVA is to be used to identify the source contributions by factors including subjects, period, formulation and potential interactions. The geometric mean ratio together with the ANOVA residual mean error term are used to identify the statistical basis for the 90% confidence interval for the ratio of the population means (Test/Reference) of the identified metrics (e.g. AUC, Cmax).
Health Canada's Part A Guide (See "Suggested Readings") provides an amplified section illustrating the calculations.

**IV.2. Testing competitive (generic) products under fed conditions**

Food-effect studies are recommended particularly for modified-release dosage forms and, in some jurisdictions, for an array of conventional solid oral products.

Commonly, aside from the incorporation of a meal, the same testing methods are to be used as described above for the fasting condition. Therefore, only the study design is presented below.

**IV.2.A. Study design**

The fed study is to be designed in such a way that the effects of formulation can be distinguished from other factors. If two formulations are being compared, a randomized two-period, two-sequence crossover study is commonly considered the design of choice. An adequate washout period between periods is needed to avoid drug carryover effects. Replicate studies, although not mandated, offer the advantage of providing a comparison of intra-subject variances for the test and reference products.

All facets of the study are to be tightly controlled. The full characteristics, including lot numbers and expiry dates, of the test and reference products shall be known. Normally, subjects fast for 10 hours prior to ingesting a standardized meal. The meal is to provide the greatest changes from the gastrointestinal physiology of a fasting state. A meal with high-fat and high-calorie content is recommended (e.g. 150, 250 and 500-600 calories from protein, carbohydrate, and fat, respectively). The meal shall be ingested over a period of 30 minutes or less. The product dose shall be ingested 30 minutes after start of the meal.

Generally, the highest safe strength/dose of the test or reference product will be administered with about 8 ounces (240 mL) of water. Further fluid shall be withheld for about 2 hours; standardized meals will be permitted beginning at four hours after drug administration. All subsequent meals will be carefully standardized.

For most drugs, subjects shall not be allowed to recline until at least two hours after product ingestion. Physical activity and posture shall be standardized to limit effects on gastrointestinal blood flow and motility. Blood samples (about 12 to 18, including a pre-dose sample) are to be drawn at appropriate, specified, and carefully recorded times (to capture increasing and decreasing concentrations during the absorption, distribution and elimination phases). The collections shall continue for about three terminal drug half-lives in order to capture at least 80% of the total area. At least three to four samples shall be obtained from the terminal log-linear phase to derive an acceptable estimate of the terminal constant ($\lambda_z$) from linear regression. For long half-life drugs, a truncated AUC (e.g. up to 72 hours) is generally considered adequate.

Blood samples or the harvested plasma/serum are to be analyzed for the administered drug or metabolites by means of a validated analytical method.

**V. REPRESENTATIVE WELL DESIGNED TRIALS**


VI. SUGGESTED READINGS

5. EMEA (European Agency for the Evaluation of Medicinal Products), CPMP (Committee for Proprietary Medicinal Products). Note for guidance on the investigation of bioavailability and bioequivalence. 2001.
Chapter 9. Pharmacoeconomics and Economic Evaluation of Drug Therapies

Tom Walley, M.D.

Professor of Clinical Pharmacology
Department of Pharmacology and Therapeutics
University of Liverpool
70 Pembrooke Place
Liverpool L69 3BX
UK
Tel 0151 794 8123
Email twalley@liv.ac.uk
I. INTRODUCTORY REMARKS

Health care funders (governments, social security funds, insurance companies) are struggling to meet their rising costs. They make many efforts to contain drug costs, by price negotiation, patient co-payments or dedicated drug budgets. Expenditure on drug therapy is a particular target for their attention for several reasons: the size of the drug bill (10-15% of most national health care budgets, and usually the second largest item after salaries); the ease of measurement of pharmaceutical costs in isolation, in contrast to most other health care costs; evidence of wasteful prescribing; and a perception that many drugs are overpriced and that the profits of the pharmaceutical industry are excessive.

But this focuses on drug costs in isolation, when what should be of greater concern to decision makers, health care professionals and the public is the value of drug therapy, a function of its benefits as well as its costs (1). Payers acknowledge that spending on drugs, which may for instance reduce the need for hospitalisation or produce greater health gain for the same resources than other medical interventions, may be a very efficient use of scarce resources. They therefore respond by demanding evidence of value for money from drug therapy. Drug therapy is open to this simply because there are high quality trials to support most new drugs’ licensing applications, in contrast to the poor evidence around most other health care interventions.

Health economics is the science of assessing cost and benefits, not to make decisions about resource use, but to inform those decisions. The aim is to identify what is most efficient, so that the greatest amount of benefit can be bought for a given amount of money or resources. But we must remember that in health care, efficiency may not be the most important objective -we might for instance prioritise caring for dying patients or treating patients with serious disease who have relatively little hope of surviving. Pharmacoeconomics is a branch of health economics that particularly considers drug therapy. It is of particular interest to pharmaceutical companies who in developing a new drug and after the traditional hurdles of efficacy, safety and tolerability must now jump over a fourth hurdle of cost effectiveness. It should also interest clinical pharmacologists, either in their roles assessing new drugs or in the conduct of clinical trials that now often include an economic component. In some areas, economic studies have become an accepted part of evaluations for reimbursement.

This chapter aims to explain the basic concepts and language of pharmacoeconomics, and of economic evaluation, and to introduce what for many clinical pharmacologists is a new area.

II. BASIC CONCEPTS AND TERMINOLOGY

Health economics is about making choices between options, when there is scarcity of resources. It is fundamentally comparative, weighing the costs and benefits of option 1 with those of option 2 (for instance, a new drug and the previous best therapy - traditional medical evaluation focused only on the benefits), to determine which is the most efficient way to use our limited resources. Efficiency is a key concept in economics, i.e. how to buy the greatest amount of benefit for a given resource use.

II.1. Opportunity cost

Another key concept is opportunity cost: this is defined as the “benefit foregone when selecting one therapy alternative over the next best alternative”. When we have limited money and we spend it on one health care intervention, we cannot spend the same money on something else. So we should be less concerned with how much a health care intervention costs, but rather with what other benefits we are giving up by using the money in that way. We need to be sure that spending money on the new therapy will buy more benefit than spending that money in some other part of the health care system.
The comparative nature of health economics means that we are interested in an incremental analysis of costs and benefits. There is usually a current treatment for most conditions, with associated costs and benefits. We would not advocate stopping all existing treatment for the condition, so the question is not what are the costs and benefits of the new treatment, but what are its added costs and benefits, over and above those of the existing treatment.

II.2. Marginal cost

A related concept is marginal costs. For instance, if a new treatment enables patients to be discharged from hospital a day earlier than an older treatment, it might be tempting to count the average cost of a hospital bed day as a saving of resources. But all the fixed capital charges for a hospital bed, which go into the average cost, e.g. costs of laboratories, kitchens, and building maintenance, will be largely unchanged. The only costs which change may be those of having a patient physically occupy the bed - the costs of the patient’s meals, treatment and perhaps nursing time. These are the marginal costs, where the resource use actually changes substantially. Incremental analysis is concerned with the marginal and not the average costs. Marginal costs are often very difficult to measure, and there is a temptation to use average costs instead. This may be justified if for instance, enough bed days are saved by the widespread adoption of a new treatment to actually reduce bed numbers and to close wards.

III. COSTS AND BENEFITS

These have broad definitions in health economics, which may depend in part on the perspective or viewpoint we choose to take. Perspective asks from whose point of view is the study conducted - from that of the health care payer, who is only interested in the direct costs of health care, or from society as a whole, where “indirect” costs (i.e. not directly on health care, such as lost of productivity etc) are also important. In general, the societal perspective is considered the most appropriate, but a health care manager with a limited budget might be tempted to ignore the societal view and consider only the costs that fall on his own budget. A study of migraine which took the health service perspective only might suggest that sumatriptan in migraine (an expensive drug in an area which previously cost the health service very little) was highly undesirable, but a study taking a societal perspective might come to the opposite conclusion (2).

III.1. Cost classification

Costs therefore can be classified as:

III.1.A. Direct – i.e. costs from the perspective of the healthcare funder: including staff costs, capital costs, drug acquisition costs. These should (in theory) be relatively easy to measure.

III.1.B. Indirect – i.e. costs from the perspective of society as a whole: for example, these might include loss of earnings, loss of productivity, loss of leisure time, due to the illness, and cost of travel to hospital etc. This would include not just the patient themselves but also their family and society as a whole. Many of these are difficult to measure, and there is some controversy over how to value these. (The UK National Institute for Clinical Excellence, NICE, adopts a limited societal perspective in its evaluations and considers the direct costs falling on the UK National Health Services, and those indirect costs funded by the state such as unemployment and sickness benefits (3)).

III.1.C. Intangible – i.e. the pain, worry or other distress which a patient or their family might suffer. These may be impossible to measure in monetary terms, but are sometimes captured in measures of quality of life.
III.2. Benefits

The benefits we expect from an intervention might be measured in:

III.2.A. “Natural” units - e.g. years of life saved, strokes prevented, peptic ulcers healed etc.

III.2.B. “Utility” units - utility is an economist’s word for satisfaction, or sense of well being, and is an attempt to evaluate the quality of a state of health, and not just its quantity. Utility estimates can be obtained through direct measurement (using techniques such as time trade off or standard gambles, or by imputing them from the literature or expert opinion. They are often informed by measures of quality of life in different disease states.

The Quality Adjusted Life Year (QALY) is one widely used measure, which attempts to integrate both quality and the quantity of life. Broadly, it assumes that if a treatment increases one’s life expectancy by 2 years, but causes adverse effects or inconvenience, such that one’s quality of life or utility are decreased by 25%, the net gain is $2 \times 0.75 = 1.5$ QALYs. QALYs are controversial for many reasons (4), not least that measuring patient utilities is difficult and preferences may change in the course of an illness (what seems an intolerable burden to a healthy individual may not seem so bad to someone who might otherwise be dead). Despite these criticisms, the concept of the QALY has advanced thinking on how to incorporate quality of life into economic evaluations.

Table 1.

<table>
<thead>
<tr>
<th>Method of economic evaluation</th>
<th>Measurement of outcome (health benefits)</th>
<th>Synthesis of costs and benefits</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cost minimisation analysis</td>
<td>Assumed to be equivalent and can take any form (e.g. number of cases detected, reductions in cholesterol levels, years of life saved)</td>
<td>Additional costs of therapy A relative to B</td>
</tr>
<tr>
<td>Cost effectiveness analysis</td>
<td>Health benefits across therapies are measured in similar natural units</td>
<td>Cost per life year gained, Cost per patient cured, Cost per life saved, etc.</td>
</tr>
<tr>
<td>Cost utility analysis</td>
<td>Health benefits across therapies are valued in similar units based on individual preferences</td>
<td>Cost per QALY gained, Cost per HYE gained</td>
</tr>
<tr>
<td>Cost benefit analysis</td>
<td>Measured in similar or different units and are always valued in monetary units (e.g., amount willing to pay to prevent a death, amount willing to pay to reduce exposure to a hazard)</td>
<td>Net benefits = Benefits minus costs, Benefit-cost ratio = benefits/costs</td>
</tr>
</tbody>
</table>

III.2.C. Associated Economic Benefit

This is usually measured in money, which is a useful common denominator allowing comparisons across different disciplines. This measure includes, for instance, the economic benefits of returning someone to work.

III.2.D. Methods of Economic Evaluation

Economic evaluation is the formal process of weighing benefits and costs in an incremental analysis. It is essentially a framework which draws up a balance sheet between costs and benefits to assist decision making.

Common Types of Study

The costs and benefits or outcome measures selected give rise to the four common types of economic evaluation (table 1). These studies are often complex and require use of economic models (a skill not dissimilar to pharmacokinetic modelling).
III.2.D.a. Cost minimisation analysis (CMA)
This involves measuring only costs, usually only to the health service, and is applicable only where the outcomes are identical and need not be considered separately. An example would be prescribing a generic preparation instead of the brand leader (lower cost but same health outcomes).

III.2.D.b. Cost effectiveness analysis (CEA)
The term cost effectiveness is often used loosely to refer to the whole of economic evaluation, but should properly refer to a particular type of evaluation, in which the health benefit can be defined and measured in natural units (e.g., years of life saved, ulcers healed) and the costs are measured in money. It therefore compares therapies with qualitatively similar outcomes in a particular therapeutic area. For instance, in severe reflux oesophagitis, we could consider the costs per patient relieved of symptoms using a proton pump inhibitor compared to those using H2 blockers. CEA is the most commonly applied form of economic analysis in the literature, and especially in drug therapy. It does not allow comparisons to be made between two totally different areas of medicine with different outcomes. The broad form of these evaluations are shown in box 1, and the key measure is the incremental cost effectiveness ratio (ICER).

Box 1.

Incremental Cost Effectiveness Ratio = \[
\frac{\text{cost of drug A} - \text{cost of drug B}}{\text{(benefits of drug A) - (benefits of drug B)}}
\]

\[
\text{ICER} = \frac{\text{difference in costs (A-B)}}{\text{difference in benefits (A-B)}}
\]

III.2.D.c. Cost utility analysis (CUA):
This is similar to cost effectiveness in that the costs are measured in money and there is a defined outcome (box 2). But here the outcome is a unit of utility (e.g., a QALY). Since this endpoint is not directly dependent on the disease state, CUA can in theory look at more than one area of medicine, e.g., cost per QALY of coronary artery bypass grafting versus cost per QALY for erythropoietin in renal disease. In practice this is not so easy since the QALY is not a well-defined fixed unit transferable from study to study. We should be particularly wary of attempts to draw up league tables of QALYs to allow comparisons between a range of therapies. The values in such tables have usually been derived at different times and in different ways and are not comparable.

Box 2. Calculating QALYs - a simple example

<table>
<thead>
<tr>
<th>With treatment X</th>
<th>Without treatment X</th>
</tr>
</thead>
<tbody>
<tr>
<td>Estimated survival = 10 years</td>
<td>Estimated survival = 5 years</td>
</tr>
<tr>
<td>Estimated quality of life (relative to ‘perfect health’) = 0.7</td>
<td>Estimated quality of life (relative to ‘perfect health’) = 0.5</td>
</tr>
<tr>
<td>QALYs = (10 X 0.7) = 7.0</td>
<td>QALYs = (5 X 0.5) = 2.5</td>
</tr>
</tbody>
</table>

QALY gain from treatment X = 7 - 2.5 = 4.5 QALYs
If the cost of treatment X is £18,000 then the cost per QALY is £4,000 per QALY (£18,000 divided between 4.5 additional QALY’s)
III.2.D.d. Cost benefit analysis (CBA)
Here, the benefit is measured as the associated economic benefit of an intervention (e.g., monetary value of returning a worker to employment earlier), and hence both costs and benefits are expressed in money. CBA may ignore many intangible but very important benefits not measurable in money terms, e.g., relief of anxiety. CBA may also seem to discriminate against those in whom a return to productive employment is unlikely, e.g., the elderly, or the unemployed.

However, the virtue of this analysis is that it may allow comparisons to be made between very different areas, and not just medical, e.g., cost benefits of expanding university education (benefits of improved education and hence productivity) compared to establishing a back pain service (enhancing productivity by returning patients to work). This approach is not widely used in health economics, although many economists like it on theoretical grounds and because it removes some of the “sacred cow” protection which surrounds health care. They argue that health should be another commodity, and not necessarily valued more than other possible uses of the resources.

III.2.D.e. Cost consequences and other types of evaluation
Other forms of quasi-health economic evaluation may be seen in the literature but are not true economic evaluations because they do not weigh costs and benefits in an incremental manner. In some cases, often where studies consider multiple outcomes, costs and benefits are presented in a disaggregated form (e.g., health profiles). These evaluations are frequently referred to as cost consequences analyses. Burden of disease (also known as cost of illness) studies attempt to measure the health and resource implications arising to society from a particular disease.

III.3. Further Points

There are two further points for definition.

III.3.A. Discounting
There is often a difference in timing between the investment of health resources and gaining the benefits. Therefore we must discount future spending, etc., to try to equalise the effects of inflation and health and financial preferences over a long period. In general, costs are discounted at an agreed rate (in the UK, currently 6% for costs). There is some debate over whether benefits can also be discounted (it is relatively easy to accept that £100 spent now is worth more than in five years time, but how does one compare a healthy year now to a healthy year in five years time?) NICE suggests discounting benefits at a rate of 1.5% (3).

III.3.B. Handling uncertainty
The measures of benefit and cost in an economic evaluation come from the medical evidence, usually clinical trials. But clinical trials address efficacy whereas health economics is more interested in effectiveness – what benefits/costs are associated with a new therapy when it is used in the real world, where patients are less well defined or monitored and where the comparator may not be the one used in the clinical trial. There is often little evidence available about effectiveness, and we are forced to make assumptions to fill the gaps in our knowledge. These assumptions should be reasonable, and should be transparent, so that they can be challenged. Any good economic study will challenge these assumptions itself, by varying them in a sensitivity analysis. This explores the extent to which a conclusion is dependent on an assumption. For instance, if a study assumes a rate of relapse of duodenal ulcers after treatment of 5% at one year, what happens if the relapse rate were to be actually 2.5%, or 10%? This might drastically affect the outcome of a study. A sensitivity analysis clarifies what are the critical assumptions and confirm that the results of the evaluation are robust, despite changes in the assumption.
IV. HANDLING THE RESULTS OF ECONOMIC EVALUATIONS

Consider the four possible results arising in a CEA (figure 1). First, if costs are lower and health benefits higher for one drug relative to another, the former is said to dominate and would be the preferred treatment (quadrant II). Second, the opposite applies, i.e. the new drug is more expensive and less effective, and thus is considered inferior and not recommended (quadrant IV). The third and most common case is where the new drug is both more effective and more expensive than the standard (quadrant I); on the basis of ICERs, a judgement must be made regarding whether the additional benefits are worth the extra costs of the new drug and, therefore, whether it is ‘cost-effective’. This might be defined by a previously agreed ICER threshold value. The fourth case is similar to the third, with the roles of the new therapy and the standard reversed (quadrant III); the question now is whether the extra benefits provided by the standard justify the additional costs of retaining it as the preferred treatment when the option of a new, cheaper but less effective drug exists.

Figure 1.

**Difference in costs**

<table>
<thead>
<tr>
<th>Quadrant</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>IV</td>
<td>The new treatment is less effective and more expensive</td>
</tr>
<tr>
<td>I</td>
<td>The new treatment is more effective and more expensive</td>
</tr>
<tr>
<td>III</td>
<td>The new treatment is less effective and less expensive</td>
</tr>
<tr>
<td>II</td>
<td>The new treatment is more effective and less expensive</td>
</tr>
</tbody>
</table>

Defining what is an acceptable maximum value or threshold for ICER is difficult and controversial, since it clearly carries an element of rationing of care. How much is an extra QALY or life year worth? This is a value judgment. It can be explored to some extent through techniques such as trying to identify what a patient or the public might be willing to pay to avoid an unfavourable outcome. There may be precedents—e.g. by common consent, we provide treatment in the form of coronary bypass grafting: we work out later that this cost £X per QALY, and so this establishes a baseline for our thinking about how much we value a QALY. In the UK, NICE seems to operate at a threshold of around £30,000 per QALY (5), although no formal threshold is declared and its existence has been formally denied. One might be more confident in setting a threshold if economic evaluations were more certain in their outcomes.

V. LIMITS OF PHARMACOECONOMIC EVALUATION

Many problems limit our use of health economics in practice (1). The whole process may be open to bias, in the choice of comparator drug, the assumptions made, or in the selective reporting of results. This suspicion arises because most studies are conducted or funded by pharmaceutical companies who
obviously are interested in the results, and there is a publication bias towards those studies favourable to sponsoring companies (6). Health economics is therefore sometimes misused as a marketing ploy. The same problems may however arise in studies funded by health care payers. To a specialist, this is not such a problem since the almost inevitable biases are usually clear. But since economic evaluation is less well understood by doctors and others, bias needs to be minimised.

Doctors may tend to equate health economics with rationing or cost cutting, and many therefore reject on principle the whole process as unethical. Since resources are limited within health services, wasting them by inefficiency is wrong, as it reduces the clinician’s ability to give the best possible care to his patients. It therefore seems unethical not to consider the economics of a medical intervention.

A key problem is our ability to implement the results of a study. No matter how good a study is, and how cost effective a therapy compared to existing treatment, it may not be possible to achieve its potential benefits because of the existing cumbersome management structures. Three problems are common: first, a short term outlook which limits the application of economic evaluations showing long term savings for the health service in return for increased spending now. Second, many budgets operate in isolation, and it is not easy to move money between them: for instance, prescribing in primary care is often funded separately from hospital services, so any increased spending on drug therapy in primary care cannot be simply funded from a future reduction in hospital admissions. Third, a new intervention may simply not be affordable no matter how cost effective it might be.

Finally, health economics and pharmaco economics is a young science and is slowly developing and testing its methodologies. We do not have space to address all of these concerns here but many of the details of the methods described above are academically and practically controversial (2). There have been many guidelines developed (e.g. ref 3) for the conduct of economic evaluation, recognising the possibilities of bias and the poor understanding of many potential users about the whole process.

**VI. THE FUTURE**

Despite these problems, economic evaluations of drug therapy are increasingly important in decision making. Clinical pharmacologists should welcome this as a means to promote efficiency and effectiveness of prescribing, and aim to move the managers’ debate away from pure cost to the question of value for money in prescribing.

**VII. FURTHER READING**

**VII.1. Useful texts**


**VII.2. Useful introductory articles for nonspecialists**

VIII. REFERENCES

Chapter 10. Drug Utilization

Folke Sjöqvist, M.D., Ph.D. 1

Donald Birkett, Ph.D. 2

1 Karolinska Institute
Division of Clinical Pharmacology
Huddinge University Hospital
S-141 86 Huddinge
SWEDEN

2 Department of Clinical Pharmacology
Flinders Medical Centre
Flinders University of South Australia
Bedford Park, SA 5042
AUSTRALIA

Parts of this chapter are extracted from the WHO booklet “Introduction to Drug Utilization Research” 2003, ISBN 92 4 156234X, with the permission of the WHO Office of Publications (Mr. D.W. Bramley). The content in this chapter represents the views of the authors and not those of WHO. The authors were main contributors to the original WHO publication.
I. INTRODUCTORY REMARKS

WHAT IS DRUG UTILIZATION RESEARCH AND WHY IS IT NEEDED

I.1. Definitions

Drug utilization research was defined by WHO in 1977 as “the marketing, distribution, prescription, and use of drugs in a society, with special emphasis on the resulting medical, social and economic consequences”. Since then, a number of other terms have come into use and it is important to understand the interrelationships of the different domains. Epidemiology is defined as “the study of the distribution and determinants of health-related states and events in the population, and the application of this study to control of health problems”. Pharmacoepidemiology applies epidemiological methods to studies of the clinical use of drugs in populations. A suitable definition of pharmacoepidemiology is: The study of the use and effects/side effects of drugs in large numbers of people with the purpose of supporting the rational and cost-effective use of drugs in the population thereby improving health outcomes.

Pharmacosurveillance and pharmacovigilance are terms used to refer to the monitoring of drug safety such as spontaneous adverse effect reporting systems, case-control and cohort studies.

Pharmacoepidemiology may be drug-oriented, emphasizing the safety and effectiveness of individual drugs or groups of drugs, or utilization-oriented aiming to improve the quality of drug therapy through pedagogic intervention. Drug utilization research may also be divided into descriptive and analytical studies. The emphasis of the former has been to describe patterns of drug utilization and to identify problems deserving more detailed studies. Analytical studies try to link drug utilization data to figures on morbidity, outcome of treatment and quality of care with the ultimate goal being to assess whether drug therapy is rational or not. Sophisticated utilization-oriented pharmacoepidemiology may focus on the drug (e.g., dose-effect and concentration-effect relationships), the prescriber (e.g., quality indices of the prescription), or the patient (e.g., selection of drug and dose vs. kidney function, drug metabolic phenotype/genotype, age, etc).

Drug utilization research is thus an essential part of pharmacoepidemiology as it describes the extent, nature and determinants of drug exposure. In common use, the distinction between these two terms has become less sharp, and they are sometimes used interchangeably. However, while drug utilization studies often employ various sources of information focusing on drugs, e.g., aggregate data from wholesale and prescription registers, the term epidemiology implies defined populations and that drug use can be expressed in terms of incidence and prevalence.

Drug utilization research and pharmacoepidemiology may provide insights into the following aspects of drug use and drug prescribing:

Pattern of use: extent and profiles of drug use and trends in drug use and costs over time.

Quality of use: audits comparing actual use to national and regional prescription guidelines or local drug formularies. Quality indices of drug use may include the choice of drug (compliance to recommended assortment), drug cost (compliance to budgetary recommendations), drug dosage (awareness of inter-individual variations in dose requirements and age dependence), drug interaction
awareness, ADR awareness, proportion of patients being aware of/unaware of the cost/benefit of the treatment, etc.

**Determinants of use:** user characteristics (e.g. socio-demographic parameters, attitude towards drugs), prescriber characteristics (e.g. specialty, education and factors influencing therapeutic decisions), and drug characteristics (e.g. therapeutic properties, affordability)

**Outcomes of use:** health outcomes (benefits and adverse effects) and economic consequences.

Pharmacoepidemiology initially focused on the safety of individual drug products (pharmacosurveillance), but now also includes studies of their beneficial effects. The driving force behind this development was a growing awareness that health outcomes of drug use in the rigorous setting of randomized clinical trials is not necessarily the same as health outcome of drug use in everyday practice. The clinical trials that are needed to obtain marketing authorization for new drugs involve limited samples of carefully selected patients, who are treated and followed-up for a relatively short period of time in strictly controlled conditions. As a result, such trials do not provide an accurate reflection of how drug use will impact health outcomes in everyday practice under everyday circumstances. Pharmacoepidemiological studies often make useful contributions to our knowledge about effectiveness and safety, because they assess drug effects in large, heterogeneous patient populations over longer periods.

Drug utilization research also provides insight into the efficiency of drug use, i.e. whether a certain drug therapy provides value for money. Drug utilization research can thus help to set priorities for the rational allocation of health care budgets.

**I.2. Why drug utilization research?**

The principal aim of drug utilization research is to facilitate rational use of drugs in populations. For the individual patient rational use of a drug implies the prescription of a well-documented drug in an optimal dose on the right indication, with the correct information and at an affordable price. Without knowledge on how drugs are being prescribed and used, it is difficult to initiate a discussion on rational drug use and to suggest measures to change prescribing habits for the better. Information on the past performance of prescribers is the linchpin of any auditing system.

Drug utilization research in itself does not necessarily provide answers, but it contributes to rational drug use in three important ways:

**I.2.A. Description of drug use patterns**

Drug utilization research will increase our understanding of how drugs are being used by:

- Making estimates of the numbers of patients exposed to drugs within a given time period. Such estimates may either refer to all drug users, regardless of when they started to use the drug (**prevalence**), or focus on patients who started to use the drug within the selected period (**incidence**).
- Describing the extent of use at a certain moment and/or in a certain area (e.g. country, region, community, hospital). Such descriptions are most meaningful when they are part of a continuous evaluation system, i.e. when the patterns are followed over time and trends in drug use can be described.
- Estimating (e.g. on the basis of epidemiological data on a disease) to what extent drugs are properly used, overused, or underused.
- Describing the pattern or profile of drug use - assessing which alternative drugs are being used for particular conditions and to what extent.
e. Comparing observed patterns of drug use with current recommendations or guidelines for the treatment of a certain disease.
f. Applying quality indicators to drug utilization patterns. An example is the so-called DU90% (drug utilization 90%), a further development of the “Top-10” list. The DU90% segment reflects the number of drugs that account for 90% of drug prescriptions and adherence to local or national prescription guidelines in this segment. This general indicator can be applied at different levels (individual prescriber, group of prescribers, hospitals, region, county, etc.) to get a rough estimate of the quality of prescribing.
g. Feeding back drug utilization data to prescribers. This is particularly useful when the individual’s drug prescribing can be compared with some form of "gold standard" or best practice, and with the average prescriptions in the country, the region, or the area.
h. Relating the number of case reports about a drug problem or adverse effects to the number of patients exposed in order to assess the potential magnitude of the problem. If it is possible to detect that the reaction is more common in a certain age group, in certain conditions or at a special dose level, improving the information on proper use such as indications, contraindications and appropriate dosages may be sufficient to assure a safer use. Thereby withdrawal of the drug from the market may be avoided.

I.2.B. Early signals of irrational use of drugs
Drug utilization research may generate hypotheses that set the agenda for further investigations by:

a. Comparing drug utilization patterns and costs between different regions or time periods. Hypotheses can be generated to form the basis for investigations of the reasons for, and health implications of, the differences found. Geographical differences and changes over time in drug use may have medical, social and economic implications both for the individual patient and for society, and are thus important to identify, explain and sometimes correct.
b. Comparing observed patterns of drug use with current recommendations/guidelines for the treatment of a certain disease. Hypotheses can then be generated about whether discrepancies represent less than optimal practice, whether pedagogic interventions (education) are required, or whether the guidelines need to be reviewed in the light of actual practice. These considerations should include both underuse and overuse of drugs.

I.2.C. Interventions to improve drug use – follow-up
Drug utilization research may enable us to assess whether interventions undertaken to improve drug use have had the desired impact by:

a. Monitoring and evaluating the effects of measures taken to improve undesirable patterns of drug use (regional or local formularies, information campaigns, regulatory policies, etc.)
b. Following the impact of regulatory changes or changes in insurance or reimbursement systems. This also requires a broad survey, because the total cost to society may remain the same or may even increase, if other more expensive drugs are used as an alternative.
c. Assessing to which extent promotional activities of the pharmaceutical industry and educational activities of the society impact on the patterns of drug use.

II. TYPES OF DRUG USE INFORMATION

Different types of drug use information are required depending on the problem being evaluated. These include information about the overall drug use, or use of drug groups, individual generic compounds or specific products. Often, information about the condition being treated, about the patient and about the
prescriber will be required. In addition, data on drug costs will be important in ensuring that drugs are used efficiently and economically. These types of drug information are described below.

II.1 Drug based information

The trends in total drug use may sometimes be useful to know, but more detailed information is usually required to answer clinically important questions. This may involve aggregation of drug use at various levels, and information on indications, doses and dosage regimens.

II.2. Problem or encounter-based information

Instead of asking how a particular group of drugs is used, one may well address the question how a particular problem (e.g. sore throat, hypertension, gastric ulcer, depression) is managed.

II.3. Patient information

Demographic and other information about the patient will often be useful. The age distribution of patients will sometimes be of critical importance, for example to assess the likelihood of severe adverse effects with NSAIDs, or whether the drug is being used in an age group different to that in which the clinical trials were performed. The co-morbidities of the patient group may be important in determining treatment choice and adverse effects. As an example in the management of hypertension, beta-blockers should be avoided in patients with asthma, and ACE inhibitors preferred in patients with heart failure.

Qualitative information such as knowledge, beliefs, and perceptions among patients and their attitudes to drugs will be important in some cases, for example in assessing patient pressures on doctors to prescribe antibiotics, or in designing consumer information/education programs.

II.4. Prescriber information

The prescriber is a critical point in determining drug use. Some sceptics even claim that doctors differ more than patients and that differences in drug prescribing often lack rational explanations. Dissecting the factors that determine prescribing behaviour is therefore often central to understanding how and why drugs are prescribed.

III. SOURCES OF DRUG UTILIZATION DATA

The drug use chain includes the processes of drug acquisition, storage, distribution, prescribing, patient compliance and review of outcome of treatment. Each of these events is an important aspect of drug utilization. Drug utilization data may be derived from quantitative or qualitative studies. Quantitative data may be used to describe the present state, and trends in drug prescribing and drug use at various levels of the health care system. Quantitative data are usually obtained from routinely collected data or from surveys. Qualitative studies assess the appropriateness of drug utilization and generally link prescribing data to reasons (indications) for prescribing. Such studies have been referred to as Drug Utilization Review or Drug Utilization Evaluation. The process is one of a “therapeutic audit” based on defined criteria and has the purpose of improving the quality of therapeutic care. There is an increasing interest in the evaluation of the economic impact of clinical care and medical technology. This has evolved into a discipline dedicated to the study of how pharmacotherapeutic methods influence resource utilization in health – pharmacoconomics.

The increasing interest in efficient use of health care resources has resulted in the establishment of computer databases for studies on drug utilization. Some of the databases can generate statistics for
patterns of drug utilization and adverse drug reactions. Data may be in the form of drug sales, drug movement at various levels of the drug distribution chain, pharmaceutical and medical billing data or samples of prescriptions. Data may also be obtained from drug importers, wholesalers or local manufacturers.

Data from medical practices and health facilities may be used to measure specific aspects of health provision and drug use. Such data may be used to generate indicators that provide information on prescribing habits and aspects of patient care. These indicators can be used to determine where drug use problems exist, provide a mechanism for monitoring and supervision and motivate health care providers to follow established health care standards.

Prescription and dispensing data are useful for determining some of the quality indicators of drug use recommended by WHO. These include:

a. Average number of drugs per prescription (encounter)
b. Percentage of drugs prescribed by generic name
c. Percentage of encounters with an antibiotic prescribed
d. Percentage of encounters with an injection prescribed
e. Percentage of drugs prescribed from essential drugs list or formulary
f. Average drug cost per encounter

III.1 Drug use evaluation

Drug use evaluation is a system of ongoing, systematic, criteria-based drug evaluation that ensures the appropriate use of drugs. Drug use evaluation is sometimes referred to as drug utilization review. It is a method of obtaining information to identify problems of drug use. Properly developed, it not only provides a means of identifying drug use problems but also provides a means to correct the problem and thereby contributes to rational drug therapy.

Drug use evaluation can assess the actual process of medication administration or dispensing (appropriate indications, drug selection, dose, route of administration, duration of treatment, drug interactions) and also assess outcomes of treatment (cured disease conditions, decreased levels of a clinical parameter). The objectives of drug use evaluation include:

a. Ensuring that drug therapy meets current standards of care
b. Controlling drug cost
c. Preventing medication related problems
d. Evaluating the effectiveness of drug therapy
e. Identification of areas of practice that require further education of practitioners

Identification of problems to be subjected to drug use evaluation may be obtained from any of the data from the practice setting section (prescription indicators, dispensing data, aggregate data). The main source of data for drug use evaluation is the patient records. An identifiable authoritative group, like the Drugs and Therapeutic Committee, usually carries out drug use reviews in the hospital or health facility. This group has the responsibility of drawing up the guidelines, criteria, indicators and thresholds for the evaluation. Drug use evaluation may be based on data collected prospectively (as drug is being dispensed or administered) or retrospectively (based on chart reviews or other data sources).

IV. DRUG CLASSIFICATION SYSTEMS

A drug classification system represents a common language for describing the drug assortment in a country or region and is a prerequisite for national and international comparisons of drug utilization
data, which have to be collected and aggregated in a uniform way. Access to standardised and validated information on drug use is essential to allow audits of patterns of drug utilization, to identify problems in drug use, to initiate educational or other interventions and to monitor the outcomes of the interventions. The main purpose of having an international standard is to be able to compare data between countries. A recent example is the international focus on creating comparable monitoring systems for cross-national antibacterial utilization patterns in the work against bacterial resistance.

**IV.1 Different classification systems**

Drugs can be classified in different ways: according to their mode of action, according to indications, or according to chemical structure. Each classification system will have its advantages and limitations and the usefulness will depend on the purpose, the setting used, and the user’s knowledge of the methodology. Comparisons between countries may require a different classification system than a local comparison (e.g. between different wards in a hospital). Of the various systems proposed over the years, only two have survived to attain a dominant position in drug utilization research worldwide. These are the Anatomical Therapeutic Classification (AT) developed by the European Pharmaceutical Market Research Association (EPhMRA) and the Anatomical Therapeutic Chemical (ATC) classification developed by Norwegian researchers. These systems were originally based on the same main principles. In the EPhMRA system, drugs are classified in groups at three or four different levels. The ATC classification system is modified and extended from the EPhMRA system by the addition of a therapeutic/pharmacological/chemical subgroup as the fourth level and the chemical substance as the fifth level.

ATC is also the basis for the classification of adverse drug reactions used by the WHO Collaborating Centre for International Drug Monitoring in Uppsala (www.who-umc.org).

The main purpose of the ATC classification is as a tool for presenting drug utilization statistics and it is recommended by the WHO to be used in international comparisons. The EPhMRA classification system is used worldwide by IMS (Intercontinental Medical Statistics) for providing marketing research statistics to the pharmaceutical industry. It should be emphasised that there are many technical differences between the EPhMRA classification and the ATC classification. Therefore, data prepared using the ATC classification cannot be directly compared with those obtained with the EPhMRA system. In 1996, WHO established the ATC/DDD system as an international standard in drug utilization studies.

**IV.2. The ATC classification system**

In the Anatomical Therapeutic Chemical classification system the drugs are divided into different groups according to the organ or system on which they act and their chemical, pharmacological and therapeutic properties.

Drugs are classified in groups at five different levels. The drugs are divided into fourteen main groups (1st level), with two therapeutic/pharmacological subgroups (2nd and 3rd levels). The 4th level is a therapeutic/pharmacological/chemical subgroup and the 5th level is the chemical substance. The 2nd, 3rd and 4th levels are often used to identify pharmacological subgroups when this is considered to be more appropriate than therapeutic or chemical subgroups.

The complete classification of glibenclamide illustrates the structure of the code:
Thus, in the ATC system all plain glibenclamide preparations are given the code A10B B01.

Alterations in the ATC classification are made when the main use of a drug has clearly changed, and when new groups are required to accommodate new substances or to achieve better specificity in the groupings.

In the ATC system drugs are separated into groups at five different levels (described above). By use of this classification system, statistics of drug utilization grouped at five different levels can be provided; from figures showing total drug use of all products classified e.g. in a main group (1st level), to figures for the different subgroups (2nd, 3rd and 4th level) and down to figures showing use of the separate substances.

The publication *Guidelines for ATC Classification and DDD Assignment* gives further and detailed information about the ATC classification. (WHO Collaborating Centre for Drug Statistics Methodology, 2003; www.whocc.no)

**V. DRUG UTILIZATION METRICS AND THEIR APPLICATIONS**

**V.1. The concept of the defined daily dose (DDD)**

The DDD is the assumed average maintenance dose per day for a drug used for its main indication in adults.

The defined daily dose is a unit of measurement and does not necessarily agree with the recommended or prescribed daily dose (PDD). Doses for individual patients and patient groups will often differ from the DDD and have to be based on individual characteristics (e.g. age and weight) as well as pharmacokinetic and pharmacogenetic considerations.

The DDD is often a compromise based on a review of the available information about doses used in various countries. The DDD may even be a dose that is rarely prescribed, because it is an average of two or more commonly used dose sizes.

Drug utilization figures should preferably be presented as numbers of DDDs/1000 inhabitants/day or, when in-hospital drug use is considered, as DDDs per 100 bed days. For antiinfectives (or other drugs normally used in short periods), it is often considered most appropriate to present the figures as numbers of DDDs per inhabitant per year.
These terms are explained in the following:

**V.1.A. DDDs/1000 inhabitants/day**
Sales or prescription data presented in DDD/1000 inhabitants/day may provide a rough estimate of the proportion of the study population that may be treated daily with certain drugs. As an example, the figure 10 DDDs/1000 inhabitants/day indicates that 1% of the population on average might get a certain drug or group of drugs every day. This estimate is most useful for chronically used drugs when there is good agreement between the average prescribed daily dose (see below) and the DDD. It may also be important to consider the size of the population used as a denominator. Usually the general utilization is calculated for the total population including all age groups. Some drug groups have very limited use among young people, with most users above the age of 45. To correct for utilization differences due to differing age structures between countries, simple age adjustments can be made by using the number of inhabitants in the relevant age group as a denominator.

**V.1.B. DDDs per 100 bed days**
This unit may be applied when in-hospital drug use is considered. As an example, 70 DDD/100 bed days of hypnotics provide an estimate of the therapeutic intensity and suggests that 70% of the in-patients might receive a DDD of a hypnotic every day. This unit is quite useful for benchmarking in hospitals.

**V.1.C. DDDs per inhabitant per year**
This term may give an estimate of the number of days for which each inhabitant is, on average, treated annually. For example, 5 DDDs/inhabitant/year indicates that the utilization is equivalent to the treatment of every inhabitant with a 5 days course during a certain year. Alternatively, if the standard treatment period is known, the total number of DDDs can be calculated as the number of treatment courses, and the number of treatment courses can then be related to the total population.

**V.2. Prescribed daily dose/Consumed daily dose**
The prescribed daily dose (PDD) is defined as the average dose prescribed according to a representative sample of prescriptions. The PDD can be determined from prescription studies and medical- or pharmacy records. It is important to relate the PDD to the diagnosis for which the dosage is based. The PDD will give the average daily amount of a drug that is actually prescribed. When there is a substantial discrepancy between the PDD and the defined daily dose (DDD), it is important to take this into consideration when evaluating and interpreting drug utilization figures, particularly in terms of morbidity.

The PDD can vary according to both the illness treated and national therapeutic traditions. There are also substantial differences between PDDs in various countries. PDDs in Asian populations are often lower than in Caucasians. The fact that PDDs may differ from one country to another should always be considered in international comparisons.

It should be noted that the prescribed daily dose does not necessarily reflect actual drug utilization. Some prescribed medications are not dispensed, and the patient does not always take all the medications that are dispensed. Specially designed studies including patient interviews are required to measure actual drug intake at the patient level (i.e. consumed daily dose).

**V.3. Volume**
Common physical units (e.g. grams, kilos, litres), numbers of packages or tablets and numbers of prescriptions are also used for quantifying drug utilization. These units can be applied only when the use of one drug or of well-defined products is evaluated.
V.4. Cost

Drug use can be expressed in terms of costs (e.g. national currency). Cost figures are suitable for an overall cost analysis of drug expenditure. International comparisons based on cost parameters can be misleading and of limited value in the evaluation of drug use. Price differences between alternative preparations and different national cost levels make the evaluation difficult. Long-term studies are also difficult due to fluctuations in currency and changes in prices. When cost data are used, an increase in the use of cheaper drugs may have little influence on the total level, while a shift to more expensive drugs is more readily noticed.

VI. SUGGESTED READINGS

Pharmacological Research in Cardiovascular Disorders

Chapter 11. Hypertensive Vascular Disease

Chapter 12. Lipid Lowering Agents

Chapter 13. Anti-atherosclerotic Drugs

Chapter 14. Heart Failure

Chapter 15. Arrhythmias
Chapter 11. Hypertensive Vascular Disease

Pierre Larochelle, M.D., Ph.D. ¹

Maxime Lamarre-Cliche, M.D. ²

¹ Directeur de la recherche clinique
Unité de recherche en pharmacologie cardiovasculaire
Institut de recherche clinique de Montréal
110, avenue des Pins Ouest
Montréal, Québec H2W 1R7
CANADA

² Chargé de recherche
Unité de recherche en pharmacologie cardiovasculaire
Institut de recherche clinique de Montréal
110, avenue des Pins Ouest
Montréal, Québec H2W 1R7
CANADA
I. INTRODUCTORY REMARKS

Over the last 3 decades, the treatment of essential hypertension has evolved with new objectives and new classes of drugs. The objectives of the treatment have changed from the reduction in blood pressure to the prevention of cardiovascular disease by the reduction and normalization of blood pressure. Treatment is also evolving towards a more individual approach of the treatment taking into considerations such variables such as age, gender, race and the associated diseases such as among others diabetes, renal function, and cardiac function. There have been new classes of drugs which have been evaluated such as the calcium channel blockers, the ACE inhibitors, the renin-angiotensin blockers and others which are currently being tested such as the aldosterone receptor blockers.

The inhibition of the renin-angiotensin system was successfully used in the treatment of hypertension and heart failure. The first approach used was the inhibition of the angiotensin-converting enzyme (ACE). However, ACE-inhibitors induce the accumulation of other peptides such as substance P or bradykinin, and consequently untoward drug effects like cough and angioedema can become apparent. A more recent approach to counteract increased blood pressure and sympathetic tone, as well as harmful cardiovascular hypertrophy and renal lesions, was to decrease the activity of angiotensin II receptors. Pharmacological blockade of AT₁-subtype angiotensin II receptors appears to be clinically equally effective but the generation of angiotensin II remains unopposed during AT₁-blockade and leaves the potential for stimulation of other angiotensin II receptor subtypes. These two classes of drugs have been proven to be effective and very well tolerated. These drugs have not been shown to be superior to conventional treatment such as diuretics in a general population but have been shown to be superior to beta blockers in some groups of patients with hypertension such as the patients with LVH (left ventricular hypertrophy). Despite the addition of these two classes of drugs, there is still definitely a need for more effective, as safe and possibly more specific drugs to treat hypertension. Moreover, the effect of antihypertensive therapy on cardiovascular events, on cerebrovascular events, on diabetic complications progression needs to be part of the evaluation process. Further studies are needed to document the long-term benefit of antihypertensive therapy alone or in association with other cardiovascular medications.

II. PHASE II STUDIES TO EVALUATE THE EFFECTIVENESS AND SAFETY OF NEW ANTIHYPERTENSIVE DRUGS

Studies will have to assess the effect of a drug directly on blood pressure, e.g. looking primarily for an antihypertensive action and they will have also to assess the effect of the drug in special situations such as renal dysfunction or others and in special populations.

Long term studies may have eventually as an objective to assess the effect of an antihypertensive agent on cardiovascular events such as coronary events, including myocardial infarction, sudden or rapid death from cardiac causes, heart failure, acute occlusion of a major feeding artery in any vascular bed other than cerebral or coronary, death from non-coronary cardiac causes, dissecting or ruptured aortic aneurysm, or death from vascular causes, cerebrovascular events, including stroke and transient ischemic attacks, and on progression of hypertensive kidney disease, diabetic nephropathy, diabetic retinopathy, and arterial wall thickness. These studies usually are conducted over the period of 5 years and are considered long-term studies.

II.1. Outline of a potential development plan

Multicenter, randomized, double-blind, placebo-controlled, parallel group study in patients with mild-to-moderate essential hypertension (WHO classification grades 1 and 2) who have been completely withdrawn from their previous antihypertensive medication or in patients who have been newly diagnosed with mild-to-moderate essential hypertension and who are not currently taking any antihypertensive
medication(s). The objective of these phase II studies is to find the effective dose of the compound compared to a placebo. These are normally dose finding placebo controlled studies. A reference drug can be included in some studies. These studies are of primary importance to determine the dose which will be used in the phase III studies to demonstrate better efficacy and tolerability than the marketed compounds.

In these phase II protocols or in phase III protocols, the evaluation of the antihypertensive drugs in special situations can be assessed. These studies need to give the information of the effect of the drug in the elderly whether defined as over 65 or 75 years. Information needs to be obtained in both groups. More and more information will be asked in the individuals over the age of 85 years. There is limited data on the benefit of treatment in this age category but we need also at least to obtain information on the effectiveness of the drug in terms of BP lowering and dose efficacy.

Protocols need to be done in patients with isolated systolic hypertension or primarily systolic hypertension. These patients are generally elderly individuals with a specific form of hypertension which is becoming more prevalent as our population ages.

Information is required on the use of the drug according to gender and race. Black population may have a different response than other races as has been shown now in various studies.

Specific information needs also to be obtained in subgroups of hypertensive patients. Those subgroups are those with renal dysfunction that is, with a creatinine clearance of <60 ml/min, <30 ml/min and lower, as well as the patients with hepatic dysfunction

Special populations which need to be studied are also those with associated disease such as diabetes and left ventricular hypertrophy. Although not a requirement, this information will become necessary to the proper utilization of the drug

II.2. Short term studies

II.2.A. Study Objectives
Primary objectives
a. To determine the efficacy of the investigational drug X at given doses compared to placebo or the reference drug Y in patients with WHO classification grades 1 and 2 uncomplicated diastolic essential hypertension (mean sitting diastolic blood pressure [MSDBP] ≥ 95 mmHg and < 110 mmHg).
b. To determine the safety of the investigational drug X compared to the placebo in patients with WHO classification grades 1 and 2 uncomplicated diastolic essential hypertension (MSDBP ≥ 95 mmHg and < 110 mmHg).

Secondary objectives
a. To determine the efficacy and safety of different doses of the investigational drug X compared to placebo in the treatment of patients with WHO classification grades 1 and 2 uncomplicated systolic essential hypertension (systolic blood pressure [MSSBP] ≥ 145 mmHg and < 180 mmHg).
b. To assess the effects of the investigational drug X and the reference drug on standing blood pressure, sitting pulse and standing pulse.

II.2.B. Primary endpoints
a. Change from baseline in MSDBP at trough.
• Response rate: patients were defined as "normalized" responders if their blood pressure values were $< 140/90$ mm Hg, or as "non-normalized" responders if the reduction in blood pressure was $10$ mm Hg or more, compared with baseline.

b. Changes in sitting blood pressure at the end of the study.

II.2.C. Secondary endpoints
a. Change from baseline in MSSBP at trough.
b. Other variables to be analyzed include the change in standing diastolic and systolic blood pressures, sitting and standing pulse.

A variable such as the ambulatory blood pressure measurement (ABPM) is a requirement now to more precisely measure the course of action of the medication over the period of $24$ hours including the peak effect and the trough effect. It will help determine the duration of action, the peak to trough ratio. Specific protocols will be planned to evaluate the effects of the drug on ABPM and some of its variables (day pressure, night pressure, smoothness of curve etc)

II.2.D. Study Design
Eligible patients enter a washout period during which antihypertensive medication is withdrawn and no other is allowed. The washout period is followed by a $2$ to $4$ -week single-blind placebo run-in period. Patients who meet the study inclusion/exclusion criteria at the end of the single-blind placebo run-in period are then randomized in double-blind fashion to either the investigational drug X or placebo once daily for an $8$-week treatment period in a parallel designed trial. The effect of the investigated medication can also be evaluated at the end of the trial by a period of drug withdrawal by recording blood pressure and adverse events from one to seven days or more after the last dose of study medication.

Blood pressure measurement
Blood pressure will be measured using a calibrated standard mercury sphygmomanometer or a calibrated electronic automatic sphygmomanometer with digital reading. Personnel recording the blood pressure should receive training on the measurement of blood pressure and follow guidelines such as the one reported in the annex 1. The standard measurement is done in the sitting position, but standing pressure is also recorded in most protocols, as well as pulse rate. Sitting and standing blood pressure will be measured and recorded at each visit. Every effort will be made to have the same staff member obtain blood pressure measurements for the same patient, at the same time of day, using the same equipment, at each visit.

Pulse rate
At each visit, the pulse rate will be measured for $30$ seconds just prior to the blood pressure measurements; once in the sitting position and once in the standing position.

Concomitant therapy
Use of the following medications may interfere with the evaluation of efficacy, safety and/or tolerability. Therefore, these medications are excluded throughout the trial, from the beginning of the washout period until the end of the double-blind treatment period. Patients who are receiving such medications should be excluded, or if ethically justified, the medications may be withdrawn according to the manufacturer's instructions.

a. Drugs approved for the treatment of hypertension even if prescribed for another indication.
b. Any antidepressant drugs in the class of MAO inhibitors and tricyclics. Other psychotropic drugs such as benzodiazepines and selective serotonin reuptake inhibitors (SSRIs) will be allowed if well tolerated when previously taken, and if the dosage is expected to remain stable throughout the study.
c. Any systemic use of corticosteroids. Topical and nasal steroid preparations may be used as needed, but not on a daily basis.

d. Use of hormonal contraceptives, including subdermal contraceptive implants, within one month prior to Visit 1 (Week -4).

e. Thyroid medication and/or oestrogen replacement therapy, unless these have been stable maintenance replacement doses for 6 months preceding Visit 1 (Week -4).

f. Insulin.

g. Chronic administration of sympathomimetic drugs such as those found in nasal decongestants (pseudoephedrine and phenylpropanolamine) and bronchodilators (e.g., metaproterenol).

h. Antacids in amounts greater than package labelling.

i. Ergot preparations.

j. Antiarrhythmic drugs; digoxin will be permitted provided serum levels have been stable and no dose adjustments have been made during the 6 months preceding Visit 1 (Week -4).

k. Diuretics of any kind.

l. Psychotropic drugs, except for hypnotics and mild anxiolytic agents such as benzodiazepines, if these were used occasionally (pm) before the start of the study.

m. Aspirin above 325 mg daily. Aspirin will be allowed at a maximum daily dose of 325 mg for cardiac protection. Patients must have been on a stable dose prior to entry and maintain the same dose throughout the study.

n. The use of drug(s) in the 6 months prior to Visit 1 (Week -4) which are potentially hepatotoxic (e.g., methotrexate) or nephrotoxic (e.g., gentamicin).

o. Antianginal medication of any kind including calcium channel blockers or beta blockers (including beta blocker eye drops).

p. Drugs which interfere with the metabolism of the other compound such as through the inhibition or stimulation of isozymes of the cytochrome P450 known to influence the compound under study should be excluded.

II.2.E. Planned sample

The sample size is calculated on the basis of the objective to be obtained. Sample size to obtain a clinically significant reduction or difference in blood pressure can be calculated with sufficient power for example. Statistical help is required for the best accurate determination of these calculations and different tools are available on the internet to calculate sample size:

a. To conclude non-inferiority of the investigational drug within a margin of 2 mm Hg when there is no true treatment difference between the investigational drug X mg and the reference drug Y mg OD.

b. To detect a true 3 mm Hg difference between the reference drug Y mg (or the investigational drug X mg) OD and placebo under the null hypothesis that the mean difference is 0.

c. To have sufficient power to examine dose-response relationship of the investigational drug at different dose levels.

II.2.F. Study population

Patients with mild-to-moderate essential hypertension (WHO classification grades 1 and 2) who have been completely withdrawn from their previous antihypertensive medication or in patients who have been newly diagnosed with mild-to-moderate essential hypertension and who are not currently taking any antihypertensive medication(s).

II.2.G. Inclusion criteria

a. Outpatients 21 and older. The age limit will be determined by the protocol or the planned development program which should include at one point studies in elderly and very elderly patients.

b. Male or female patients are eligible. Female patients must be either post-menopausal for one year or surgically sterile, or using effective contraceptive methods such as barrier method with
spermicide or an intra-uterine device. Oral or implant contraceptives are not be allowed because of their interactions and effect on blood pressure.

c. Patients with mild to moderate essential diastolic hypertension (WHO classification grades 1 or 2) measured by calibrated standard sphygmomanometer or calibrated standardized automatic sphygmomanometer. Patients must have a MSDBP ≥ 95 mmHg and < 110 mm Hg. Blood pressure criteria will be different in studies such as those in diabetic patients.

d. Patients must have a variability of ≤ 10 mmHg in their MSDBP between pre-randomization.

e. Patients who are eligible and able to participate in the study, and who consent to do so after the purpose and nature of the investigation have been clearly explained to them (written informed consent).

II.2.H. Exclusion criteria
Patients with any of the following physiological states or concomitant medical conditions will be excluded from further participation in the study.

a. Severe hypertension (defined as MSDBP ≥ 110 mmHg and/or MSSBP ≥ 180 mmHg or 200 mm Hg).
b. Inability to discontinue all prior anti-hypertensive medications safely for a period of 12 weeks.
c. Known Keith-Wagener grade III or IV hypertensive retinopathy.
d. History of hypertensive encephalopathy or cerebrovascular accident within the preceding 6 months.
e. Transient ischemic cerebral attack during the 12 months prior to Visit 1 (Week -4).
f. Evidence of a secondary form of hypertension, such as coarctation of the aorta, hyperaldosteronism, unilateral renal disease, or pheochromocytoma, etc.
g. Type 1 diabetes mellitus.
h. Type 2 diabetes mellitus with poor glucose control as defined by fasting glycosylated hemoglobin (HbA1c) >8% or requiring insulin treatment.
i. Known or suspected contraindications, including history of hypersensitivity to the class of antihypertensive agent.
j. Any surgical or medical condition which might significantly alter the absorption, distribution, metabolism, or excretion of any drug including but not limited to any of the following:
   • History of major gastrointestinal tract surgery such as gastrectomy, gastroenterostomy, or bowel resection.
   • Currently active or previously active inflammatory bowel syndrome during the 12 months prior to Visit 1 (Week -4).
   • Currently active gastritis, duodenal or gastric ulcers, or gastrointestinal/rectal bleeding during the 3 months prior to Visit 1 (Week-4).
   • Any history of pancreatic injury, pancreatitis or evidence of impaired pancreatic function/injury as indicated by abnormal lipase or amylase.
   • Evidence of hepatic disease as determined by any one of the following: SGOT or SGPT values exceeding 2 x ULN at Visit 1 (Week -4), a history of hepatic encephalopathy, a history of oesophageal varices, or a history of portocaval shunt.
   • Evidence of renal impairment as determined by any one of the following: serum creatinine > 1.5 ULN or a level higher than normal depending on the population under study, a history of dialysis, or a history of nephrotic syndrome.
   • Current obstruction of the urinary tract or difficulty in voiding due to mechanical or inflammatory conditions which is likely to require intervention during the course of the study or is regarded as clinically meaningful by the investigator.
k. History or diagnosis of heart failure during the 6 months prior to Visit 1 (Week -4).
l. History of myocardial infarction during the 6 months prior to Visit 1 (Week -4).
m. Second or third degree heart block without a pacemaker.
n. Unstable angina pectoris.
o. Concurrent potentially life threatening arrhythmia or symptomatic arrhythmia.
p. Clinically significant valvular heart disease.
q. Volume depletion or dehydration.
r. History of malignancy including leukemia and lymphoma (but not basal cell skin cancer) within the past five years.
s. History of any severe or life-threatening disease(s).
t. Any surgical or medical condition, which in the opinion of the investigator, may place the patient at higher risk from his/her participation in the study, or is likely to prevent the patient from complying with the requirements of the study or completing the trial period.
u. History of drug or alcohol abuse within the last 12 months.
v. History of non-compliance to medical regimens or unwillingness to comply with a study protocol.
w. Participation in any investigational drug trial within one month of Visit 1 (Week -4).
x. Unwillingness or inability to give informed consent.
y. Persons directly involved in the execution of this protocol.
z. Pregnant or nursing women or women of childbearing potential not practicing effective contraceptive methods.
aa. History of clinically significant allergies, including asthma, or multiple drug allergies.
bb. History of autoimmune disorders, such as, but not limited to rheumatoid arthritis, systemic lupus erythematosus or glomerulonephritis.

II.2.I. Tools for assessing endpoints

Efficacy assessment
Using a calibrated standard sphygmomanometer and appropriate size cuff, arterial blood pressure determinations will be made in accordance with the Seventh report of the Joint National Committee on Prevention, Detection, Evaluation, and Treatment of High Blood Pressure or that of other societies (1). With the arm supported at the level of the heart, systolic pressure will be recorded when the initial sound is heard (Phase I of the Korotkoff sound); diastolic pressure will be recorded at the disappearance of the sound (Phase V of the Korotkoff sound). The cuff should be deflated at a rate not greater than 2 mm Hg/sec.

The ambulatory blood pressure measurement (ABPM) will be measured using standard procedures (Appendix 2).

Safety assessments
Safety assessments will consist of monitoring and recording all adverse events (AEs) and serious adverse events (SAEs), the regular monitoring of haematology, blood chemistry and urine values, regular measurement of vital signs and the performance of physical examinations. An ECG evaluation will be obtained at Visit 2 (Week -2).

II.2.J. Specific criteria for early withdrawal and discontinuation
a. Reasons why a patient may discontinue participation in a study include the following:
b. Adverse event(s)
c. Abnormal laboratory value(s)
d. Abnormal test procedure result(s)
e. Unsatisfactory therapeutic effect
f. Patient's condition no longer requires study treatment
g. Protocol violation
h. Subject withdrew consent
i. Lost to follow-up
j. Administrative reasons
k. Death
Patients with MSDBP $\geq 110$ mmHg or MSSBP $\geq 180$ mmHg at any time during the single-blind or double-blind treatment phases must be permanently discontinued from the trial.

II.2.K. Data analysis method
The proportion of patients in each treatment achieving a successful reduction in MSDBP during the double-blind period will be compared at the endpoint using a one-way logistic model with treatment as the factor at Visit 7 (Week 8) for all randomized patients. Success is defined as a mean sitting diastolic blood pressure $< 90$ mmHg or a decrease $\geq 10$ mmHg sitting diastolic blood pressure from the randomization visit.

The changes from baseline will be analyzed using a two-way analysis of covariance model with treatment and center as factors, and the baseline as a covariate as well as treatment-by-baseline interaction. All pairwise treatment comparisons will be made based on this analysis model. Due to a large number of study centers and treatment groups planned in this study, treatment-by-center interaction effect may be difficult to interpret in a statistical model. However, a summary of means by treatment and center will be provided for the primary analysis.

III. PHASE III STUDIES TO EVALUATE THE EFFECTIVENESS AND SAFETY OF NEW ANTIHYPERTENSIVE DRUGS

III.1. Outline of a typical development plan
The phase III studies have the objective to demonstrate and compare the efficacy of the antihypertensive agent against the usual antihypertensive compounds. The new compound will be tested for efficacy and compared to conventional treatments such as diuretics (hydrochlorothiazide), beta blockers (metoprolol or atenolol), long acting dihydropyridine calcium channel blockers (amlodipine or long acting nifedipine), ACE inhibitors (enalapril or lisinopril) and AT$_1$ receptor blockers such as losartan.

Since the treatment of hypertension requires more than one antihypertensive agent, it will also be a requirement to test the new agent in combination to the most commonly used other antihypertensive agent either as a specific trial such as a combination with a thiazide diuretic or as an add-on protocol where non responsive patients will receive add on treatments to normalize blood pressure. Combination will also be required where the new medication will be added as a second line agent to conventional treatment to evaluate potential synergistic effect on BP reduction but also potential adverse effects.

Phase III long-term studies are done to follow up on the long term effectiveness on blood pressure and adverse effects of the medication over a period which could be up to two years of follow up in small number of patients who generally have participated in phase II or III studies and are asked to continue on the medication in an open study of efficacy and safety

III.2. Short-term studies

III.2.A. Objectives
The objective is to demonstrate the relative efficacy and safety of the new compound.

III.2.B. Primary endpoints
a. The decrease in blood pressure whether systolic or diastolic blood pressure measured in standard conditions at through compared to other antihypertensive treatments.

b. Responders rate are also measured in terms of % of patients obtaining a predetermined endpoint of BP such as diastolic BP less than 90 mm Hg and systolic pressure less than 160 mm Hg or 140 mm Hg.
III.2.C. Secondary end-points
As described for phase II studies in II.2.C.

III.2.D. Study design
Eligible patients will enter a 2 to 4 weeks washout period followed by a 2 to 4 weeks single blind placebo period to determine if patients can meet the inclusion criteria for blood pressure or if they have to be excluded for other exclusion criteria. They will then be randomized to either treatments conventional or experimental to evaluate efficacy and safety. The duration of the short term active treatment phase is of the order of 8 weeks

III.2.E. Planned sample
As described for phase II studies in II.2.E.

III.2.F. Study population
As described for phase II studies in II.2.F.

III.2.G. Specific inclusion criteria
As described for phase II studies in II.2.G.

III.2.H. Specific exclusion criteria
As described for phase II studies in II.2.H.

III.2.I. Tools for assessing endpoints
As described for phase II studies in II.2.I.

III.2.J. Specific criteria for early withdrawal and discontinuation
As described for phase II studies in II.2.J.

III.2.K. Data analysis method
The proportion of patients in each treatment achieving a successful reduction in MSDBP during the double-blind period will be compared at the endpoint using a one-way logistic model with treatment as the factor at Visit 7 (Week 8) for all randomized patients. Success is defined as a mean sitting diastolic blood pressure < 90 mmHg or a decrease ≥10 mmHg sitting diastolic blood pressure from the randomization visit.

The changes from baseline will be analyzed using a two-way analysis of covariance model with treatment and center as factors, and the baseline as a covariate as well as treatment-by-baseline interaction. All pairwise treatment comparisons will be made based on this analysis model. Due to a large number of study centers and treatment groups planned in this study, treatment-by-center interaction effect may be difficult to interpret in a statistical model. However, a summary of means by treatment and center will be provided for the primary analysis.

The following pairwise comparisons shall be performed:
  a. Investigational vs. reference drug
     • Primary comparison: non-inferiority and/or superiority of the investigational drug X mg OD vs. reference drug Y mg OD.
       i. Step 1: non-inferiority Test (one-sided 97.5% CI) to show the investigational drug is as good as or not worse than the reference drug by a predefined margin (2,3).
       ii. Step 2: superiority Test (one-sided $\alpha = 0.025$) to show the investigational drug is superior to the reference drug. This test is performed only if non-inferiority in Step 1 is shown.
Secondary comparison: superiority of the investigational drug 2X and 4X mg OD vs. reference drug Y mg.

b. Investigational drug vs. placebo

Primary analysis: dose-response via a regression analysis. A second-order regression analysis with the dose as predictor variable will be performed for the change from baseline in MSDBP at Visit 7 (Week 8) and endpoint to examine the relationship between the efficacy response and the dose. A test for lack-of-fit will be performed at significance level of 0.1.

Secondary analysis: pairwise comparison of investigational drug X, 2X and 4X mg OD vs. placebo.

The statistical test for each of the pairwise comparisons will be made at a two-sided 0.05 statistical significance level. Summary statistics for the changes from baseline of efficacy variables will be presented by treatment group and time point, as well as by treatment group, trial center, and visit; treatment group, age and visit; treatment group, sex and visit; and treatment group, race, and visit. Within-treatment analysis for all the efficacy variables will be performed by a paired t-test at the endpoint.

IV. OTHER STUDIES

PHASE IV MORTALITY AND MORBIDITY STUDIES

IV.1. Outline of a typical development plan

To evaluate the effect of antihypertensive therapy on cardiovascular events in hypertensive patients; the outcomes in subjects with hypertension who were treated with the new medication will be compared with the outcomes in those treated with conventional treatments. The study will be a double-blind, multicenter, randomized, parallel-group trial in subjects with essential hypertension (sitting blood pressure 160–200/95–115 mm Hg) or has been in some occasions a PROBE study (Prospective Open Blinded Endpoints).

IV.2. Long-term studies

IV.2.A. Objectives

Primary objectives
To compare the effect of either regimens in preventing cardiovascular complications either cardiac, cerebrovascular or a combination of endpoints which generally include cardiovascular death, acute MI, stroke and heart failure requiring hospitalization.

Secondary objectives
To compare the two regimens on some individual endpoints:
  a. hospitalization for angina, cardiac revascularization, heart failure, transient ischemic attack, accelerated or malignant hypertension, or renal failure in addition to the primary outcome;
  b. all-cause mortality;
  c. cancer.

IV.2.B. Primary endpoints

IV.2.C. Secondary endpoints

IV.2.D. Study design

Patients are randomized to once daily treatments under study-based antihypertensive treatment in a parallel-group for at least 4 years and until the calculated number of patients has a primary cardiovascular
event which has been validated (death, myocardial infarction, or stroke). The patients will be followed for
the duration of period calculated to be necessary to see the predicted number of events. Patients are
generally followed for a period of 4 to 6 years depending on the number of events required. They are
followed at regular visits to obtain target blood pressures of less than 140/90 mm Hg. Central laboratory
are required to standardize measurements such as ECG, echocardiogram or other specific measurements
being used as endpoints.

As for other large morbidity and mortality trials, there are independent committees to adjudicate events
and Data Monitoring safety Boards to guarantee the security and safety of the participants. The trial is
generally run by a steering committee which determines a steering committee to organize the protocol

Other medications
Usually, other drugs are permitted if clinically indicated. The choice of the drug will depend upon the
drugs on study, in the sense that drugs of different class may be added. Any additional antihypertensive
agent may be added as a step 3 medication (non-blinded).

IV.2.E. Planned sample
If the trial aims to document the effect of antihypertensive therapy on cardiovascular morbidity and
mortality, patients with secondary hypertension, myocardial infarction or stroke within the previous 6
months, angina pectoris requiring treatment with β-blockers or calcium-antagonists, heart failure or left
ventricular ejection fraction of 40% or less, or with a disorder that, in the treating physician's opinion,
requires treatment with drugs of the same class of the tested drugs will be excluded.

The design of a phase IV study of morbidity and mortality requires the use of statisticians in the
calculation of sample sizes based on the requirements determined by the investigators. These calculations
take into account the primary endpoints, the potential duration of the study and the population under study
to obtain predetermined objectives. It also assumes a percentage of non-compliance to the study, drop-outs
to the trial and loss to follow-up during the trial. The population when comparing two active based
treatments can be in the order of 15,000 patients up to 40,000 as was seen in the ALLHAT trail recently
with a follow-up of 4 years (4).

IV.2.F. Study population

IV.2.G. Specific inclusion criteria

IV.2.H. Specific exclusion criteria

IV.2.I. Tools to assess endpoints

Efficacy assessment
The method of assessment will differ greatly depending upon the primary and secondary objectives of the
trial. Trials have assessed the effect of test drugs on cardiovascular morbidity and mortality documenting
left ventricular hypertrophy by means of electrocardiograms (5). In other trial, deaths were documented
through the National Death Index; acute MI required two out of three of the following conditions:

a. symptoms compatible with acute MI (e.g., chest pain) lasting longer than 15 minutes;
b. electrocardiographic changes (new persistent ST-segment elevation or pathological Q waves in 2
   contiguous leads); and
c. increased cardiac enzymes (more than twice the upper limit of normal).

A diagnosis of stroke required the presence of focal neurological deficit lasting longer than 24 hours.
Imaging studies were not required to document a stroke. Any death thought to be compatible with
coronary heart disease (e.g., heart failure, sudden death) or cardiovascular disease was counted as a cardiovascular disease-related death (6).

An additional trial assessed the study outcomes at follow-up visits and reported to the clinical trials center. Hospitalized outcomes were primarily based on clinic investigator reports, and copies of death certificates and hospital discharge summaries were requested. In addition, searches for outcomes were accomplished through the Center for Medicare and Medicaid Services, the Department of Veterans Affairs, the National Death Index, and the Social Security Administration databases. A death was ascertained by clinic report or by match with the aforementioned databases plus a confirmatory death certificate. Medical reviewers from the clinical trials center verified the physician-assigned diagnoses of outcomes using death certificates and hospital discharge summaries. More detailed information was collected on a random subset of CHD and stroke events to validate the procedure of using physician diagnoses (4).

The frequency of the measure of endpoints depends on the objective of the trial. For instance, to assess cardiovascular endpoints, in one trial, participants were seen at least semi-annually for blood pressure measurements, treatment dispensing, and endpoint surveillance. On-site data verification was performed at least annually. An independent data and safety monitoring board met semi-annually to review accumulating data. Confidence intervals based on the Lan-DeMets version of the O'Brien-Fleming group sequential boundaries were used as guidelines for early termination. All analyses were performed independently of the sponsor. All study investigators and the study sponsors were blinded to all between-treatment comparisons until completion of endpoint data collection and review (6).

In another trial, patients were followed for at least 4 years with regular visits and increases in drug doses to reach a target blood pressure of less than 140/90 mm Hg. All screening, baseline, serial, yearly, and endpoint electrocardiograms were centrally assessed for signs of LVH and Minnesota coded at one reading center. Since combined ECG assessment of QRS voltage and duration enhances sensitivity for detection of LVH at acceptable levels of specificity, the product of QRS duration and Cornell voltage (with adjustment of 8 mm in women and a partition value of >2440 mm×ms) was used to recognize LVH. These composite ECG criteria have about 95% specificity in healthy people and 50% sensitivity in patients with LVH ascertained at necropsy or by echocardiography LVH (5).

IV.2.J. Specific criteria for early withdrawal and discontinuation
As described for phase II studies in II.2.J.

IV.2.K. Data analysis method
In one trial, time to event methods (Cox proportional hazards model and Kaplan Meier curves) were used to compare outcomes for participants randomly assigned to the investigational drug and the comparators. Analyses were by modified intent to treat (modified by the exclusion of 2 sites with data integrity concerns), unless otherwise specified, and were stratified by the choice of standard of care and geographic region in which the participant's clinical site was located. Analyses of primary and secondary events considered censoring due to losses to follow-up (e.g., participants for whom the primary event status was unknown on the closing date), non-cardiovascular disease-related deaths (as appropriate), and the closing date of the study. Losses were censored at the date the primary event status was last known (either the date provided by the site during the closeout process; or the date of the last follow-up visit). The proportional hazards assumption was tested by including an interaction term between the randomized treatment indicator and log-transformed follow-up time. Blood pressure changes from baseline were compared between the 2 treatment groups using, the t test. All analyses were performed using SAS statistical software (Version 8.0, SAS Institute Inc, Cary, NC) (6).

In the second trial, analysis of all cardiovascular endpoints was by intention to treat; all randomized patients were included in their treatment group, and all available follow-up data were included from
randomization to the end of the study. The difference between treatment groups with respect to clinical events was assessed by a Cox regression model with degree of LVH (measured as a continuous variable) and the Framingham risk score defined by baseline characteristics as covariates. Treatment effects were measured by hazard ratios (relative risks) and 95% CIs by Cox regression models. The risk reduction for drug XXX against drug YYY was calculated as $100 \times (1 - \text{relative risk})$. Event rates over time are presented as Kaplan-Meier curves. Adjustment for blood pressure was derived from Cox regression models with blood pressures throughout the trial as time-varying covariates. Differences between groups in changes in ECG measures of LVH were analysed with the Wilcoxon rank-sum test, and the frequency of adverse experiences with Fisher's exact test (5).

V. REFERENCES


VI. EXAMPLES OF LANDMARK WELL DESIGNED TRIALS


APPENDIX 1 – Canadian Hypertension Recommendations

Recommended techniques for measuring blood pressure

1. Measurements should be taken with a sphygmomanometer that is known to be accurate. Although a mercury manometer may be preferable, a recently calibrated aneroid, or a validated and recently calibrated electronic device, can be used. Aneroid devices and mercury columns need to be clearly visible at eye level.

2. Choose a cuff with an appropriate bladder width matched to the size of the arm. The optimal bladder width equals the arm circumference/2.5, with an acceptable range of 80% to 100% of the arm circumference.

3. Place the cuff so that the lower edge is 3 cm above the elbow crease and the bladder is centred over the brachial artery. The patient should be resting comfortably for 5 min. in the seated position with back support. The arm should be bare and supported with the antecubital fossa at heart level because a lower position will result in an erroneously higher systolic blood pressure and diastolic blood pressure. There should be no talking, and the patient’s legs should not be crossed. At least two measurements should be taken in the same arm with the patient in the same position. Blood pressure also should be assessed after 2 min. of standing and at times when patients report symptoms suggestive of postural hypotension. Supine blood pressure measurements may also be helpful in assessing elderly and diabetic patients.

4. Increase the pressure rapidly to 30 mmHg above the level at which the radial pulse is extinguished (to exclude the possibility of a systolic auscultatory gap). Continue to auscultate at least 10 mmHg below phase V to exclude a diastolic auscultatory gap.

5. Place the bell or diaphragm of the stethoscope gently and steadily over the brachial artery.

6. Open the control valve so that the rate of deflation of the cuff is approximately 2 mmHg/heart beat. A cuff deflation rate of 2 mmHg/heart beat is necessary for accurate systolic and diastolic estimation.

7. Read the systolic level – the first appearance of a clear tapping sound (phase I Korotkoff) – and the diastolic level B the point at which the sounds disappear (phase V Korotkoff). Record the blood pressure to the closest 2 mmHg on the manometer (or 1 mmHg on electronic devices), as well as on the arm used, and note whether the patient was supine, sitting or standing. Record the patient’s heart rate. The seated blood pressure is used to determine and monitor treatment decisions. The standing blood pressure is used to examine for postural hypotension, if present, which may modify the treatment.

8. If Korotkoff sounds persist as the level approaches 0 mmHg, then the point of muffling of the sound is used (phase IV) to indicate the diastolic pressure.

9. In the case of arrhythmia, additional readings may be required to estimate the average systolic and diastolic pressure. Isolated extra beats should be ignored. Note the rhythm and pulse rate.

10. Leaving the cuff partially inflated for too long will fill the venous system and make the sounds difficult to hear. To avoid venous congestion, it is recommended that at least 1 min. should elapse between readings.

11. Blood pressure should be taken at least once in both arms, and if an arm has a consistently higher pressure, then that arm should be clearly noted and subsequently used for blood pressure measurement and interpretation.
Ambulatory blood pressure measurements will be made using the Spacelabs Model 90207 monitor.

On the days that the ABPM equipment will be applied, patients should arrive at approximately 8h00 AM to allow additional time for ABPM procedures such that dosing of medication occurs as close to 9:00 AM as possible.

The ABPM monitors will be programmed to measure blood pressure every 30 minutes throughout the day (0500 – 2300) and every 60 minutes at night (2300 – 0500). Patients will be advised not to move the arm during each blood pressure measurement and will also be given instructions concerning interruption of measurement in case of malfunction of the device or repositioning of the cuff if it slips.

For each of the two 24-hour ABPM monitoring sessions, the following procedures will be performed:

1. **Do not reuse batteries.** Always install four (4) fresh AA batteries prior to initializing the monitor.
2. Connect the cable from the modem to the monitor.
3. When attaching the monitor to the patient, first palpate the patient’s brachial artery (you can mark location with felt-tip pen) then apply an appropriate sized ambulatory BP cuff to patient’s non-dominant arm (e.g. if patient is right-handed, apply cuff to left arm).
4. Take up to five correlation readings. Attach T-tube to office sphygmomanometer, monitor and cuff. Allow cuff to inflate and listen to pressure. The monitor must be within 10 mm Hg of the pressure obtained by the column of mercury. If not, adjust cuff and try again.
5. Remove T-tube and attach cuff to monitor.
6. Manually trigger one or two ABPM readings to make sure the monitor is working properly. Give the patient a dose of study medication and then manually trigger another ABPM reading. The official “dose time” for the 24-hour recording will be the time shown on the ABPM clock as you pressed the blue start/stop button for the manual reading. Do not use your watch or a wall clock. Record the noted ABPM clock time on the CRF as the dose time. Dosing must occur at 8:00 AM plus or minus one hour.
7. Just prior to removal of the monitor, take a final manual reading using the start/stop button. The final manual reading should occur as close to 24 hours since prior dosing as recorded on the CRF. (The patient might need to repeat the ABPM if the monitoring is < 24 hours in duration, reference Appendix 11.3.3.) If you see “Ecxx” (i.e. error code) on the ABPM display, take another reading to ensure you have a captured BP reading. Remove the monitor from the patient’s arm.
8. After downloading of your patient’s data is complete, you may initialize the monitor for future use, making certain to first install fresh batteries. Make sure the monitor is turned off for storage.

The data collected will be evaluated to determine if it meets the criteria for a successful monitoring session.

**Operating instructions for ambulatory blood pressure monitoring**
Refer to the operating manual for a detailed explanation of ABPM operation.

**HANDLING OF ABPM DATA FOR ANALYSIS**
All data editing will be performed on blinded data with no further editing performed once patient treatment assignments are known.

**Screening Rules for Individual Readings**
The following screening rules will be used by the ABPM vendor to evaluate the validity of the individual readings from a patient’s monitor.

Screening Rules For Individual 20-Minute Interval Readings:
1. If the observed systolic blood pressure reading is either <50 mm Hg or >250 mm Hg, then the entire monitoring record will be considered invalid.

2. If the observed diastolic blood pressure reading is either <20 mm Hg or >130 mm Hg, then the entire monitoring record will be considered invalid.

3. If the calculated pulse pressure (i.e. SBP minus DBP) is either <15 mm Hg or >150 mm Hg, then the entire monitoring record will considered invalid.

4. If the observed pulse rate reading is either <20 bpm or >200 bpm, then the entire monitoring record will be considered invalid.

An entire monitoring record refers to all readings and/or calculations, e.g. SBP, DBP, mean arterial pressure (MAP), pulse pressure (PP), and pulse rate (PR), for a particular 20-minute interval.

In addition, prior to statistical analysis of the ABPM results, the following screening rule will be used to further determine the validity of the individual SBP, DBP, and PR readings from a patient’s monitor.

5. For each observed reading, the six readings surrounding the observed reading (i.e. three before and three after) will be averaged together (the average will exclude the observed reading). If the observed reading differs by more than three standard deviations of the mean of the six surrounding readings and is outside of the range of values considered plausible for that particular measurement (see below) then the observed reading will be considered invalid. For observed readings that are numbered 1, 2, 3, 3rd to last, 2nd to last, or last, a wrap-around technique will be used to apply this screening rule.

   Plausible Range of Observed Readings:

<table>
<thead>
<tr>
<th>Measurement</th>
<th>Plausible Range</th>
</tr>
</thead>
<tbody>
<tr>
<td>SBP</td>
<td>mean ± 40 mm Hg</td>
</tr>
<tr>
<td>DBP</td>
<td>mean ± 20 mm Hg</td>
</tr>
<tr>
<td>PR</td>
<td>mean ± 20 bpm</td>
</tr>
</tbody>
</table>

**Calculation of hourly means**

Hourly means relative to both the dosing time and clock time will involve only valid 20-minute interval ABPM records. No imputation of missing or invalid readings/records will be performed. Additionally, one valid 20-minute interval reading per hour will be adequate for a hourly mean for each of the variables (e.g. SBP, DBP, MAP, PP, and PR).

**Hourly means relative to dose time**

Hour relative to dose time will be defined starting at dose time and incremented every 60 minutes. Valid readings collected at dosing and up to (but not including) one hour after dosing will be averaged to yield a 1-hour post-dose mean, valid readings collected at one hour post-dose and up to (but not including) two hours post-dose will be averaged to yield a 2-hour mean, etc. Hourly means will then be calculated for each of the hours post-dose that the monitor recorded measurements.

For example, if dosing was at 08:14 then records taken at 08:14 to 09:13 will be considered as within dose time hour 1, records taken at 09:14 to 10:13 will be considered as within dose time hour 2, etc., with hourly means relative to dose time calculated accordingly.

**Hourly means relative to clock time**

Clock time hour will be defined as given below with hourly means calculated for each clock time hour.

<table>
<thead>
<tr>
<th>Time Interval</th>
<th>Clock Time Hour</th>
</tr>
</thead>
<tbody>
<tr>
<td>24:00 – 00:59</td>
<td>0</td>
</tr>
<tr>
<td>01:00 – 01:59</td>
<td>1</td>
</tr>
<tr>
<td>02:00 – 02:59</td>
<td>2</td>
</tr>
<tr>
<td>03:00 – 03:59</td>
<td>3</td>
</tr>
<tr>
<td>12:00 – 12:59</td>
<td>12</td>
</tr>
<tr>
<td>13:00 – 13:59</td>
<td>13</td>
</tr>
<tr>
<td>14:00 – 14:59</td>
<td>14</td>
</tr>
<tr>
<td>15:00 – 15:59</td>
<td>15</td>
</tr>
</tbody>
</table>
Readings of valid records taken at the end of a monitoring that are less than 24 hours since dosing and which would have the same clock time hour as the first clock time hour defined at the beginning of the dosing period will be used. All readings taken at or beyond 24-hours after dosing will not be used in the calculation of hourly means relative to clock time. For example, if dosing was at 08:14 then records taken the following morning from 08:00 to 08:13 will be used in the calculation of the hourly mean for clock time hour 8 while all records taken after 08:13 will not be used.

**Criteria for a successful monitoring**
The following rules will be used to evaluate whether the entire 24-hour interval of readings is unsuccessful and may need to be repeated.

Criteria for Successful Monitors:
The following types of monitors will **not** be considered successful:
1. those with more than a total of six non-consecutive hourly means missing during the 24-hour dosing period, or
2. those with more than three consecutive hourly means missing during the entire 24-hour dosing period.

Although there are potentially inherent differences in the hourly means relative to dose time and relative to clock time, as well as inherent differences due to the method of defining clock time hour, determination of whether or not a monitoring is “successful” and potentially needs to be repeated will be based on the ABPM vendor’s calculations.

**ABPM Derived Endpoints**
For all successful monitorings at baseline the ABPM vendor will calculate the 24-hour mean relative to clock time for diastolic blood pressure (DBP) to determine whether the patient qualifies for randomization into the active treatment phase.

Additionally, for all successful ABPMs the following endpoints as defined below will be derived:
1. Last 6-hour mean for systolic blood pressure (SBP), diastolic blood pressure (DBP), mean arterial pressure (MAP) and pulse rate (PR), based on the hourly means relative to dose time,
2. 24-hour mean for SBP, DBP, MAP and PR, based on the hourly means relative to dose time,
3. Morning (06:00-11:59) mean for SBP, DBP, MAP and PR, based on the hourly means relative to clock time,
4. Daytime (06:00-21:59) mean for SBP, DBP, MAP and PR, based on the hourly means relative to clock time,
5. Night time (22:00-05:59) mean for SBP, DBP, MAP and PR, based on the hourly means relative to clock time,
6. Systolic and diastolic load (i.e. the overall percentage of valid measurements for SBP above 140 mm Hg during the daytime and above 130 mm Hg during the night time, and the overall percentage of valid measurements for DBP above 90 mm Hg during the daytime and above 85 mm Hg during the night time, respectively).
ABPM response rate definitions are:
1) ABPM DBP "control" rate: 24-hour mean DBP < 80 mm Hg
2) ABPM DBP "response" rate: 24-hour mean DBP < 80 mm Hg or a reduction from baseline of ≥ 10 mm Hg
3) ABPM SBP "response" rate: 24-hour mean SBP < 130 mm Hg or a reduction from baseline of ≥ 10 mm Hg.

**Graphical Presentation of Hourly Means**
The hourly ABPM means (relative to dosing time) at baseline and at the end of the study will be averaged over all patients to get overall mean blood pressure profiles over the 24 hour post-dose period. These mean profiles will be graphically displayed for all treatment groups.

The mean profiles of the change from baseline in hourly means (relative to dose time) will also be graphically displayed for all treatment groups.

It is only intended to perform these analyses for the primary ABPM analysis dataset as specified in the STATISTICS section of the protocol.
Chapter 12. Lipid Lowering Agents

Robert Dufour, M.D.

Directeur associé
Unité de recherche en hyperlipidemies et athérosclerose
Institut de recherches cliniques de Montréal
120, avenue des Pins Ouest
Montréal, Québec H2W 1R7
CANADA
Tel 514 987-5758
Email robert.dufour@ircm.qc.ca
I. INTRODUCTORY REMARKS

Atherosclerotic cardiovascular disease is still the number one killer and the main cause of morbidity in the world. The link between high blood cholesterol, atherosclerosis and coronary heart disease (CHD) has been known for decades, but the scientific community has paid attention to this major risk factor only since powerful lipid lowering agents (LLA) were developed and commercialized about 15 years ago. The new statins (HMG Co A reductase inhibitors) can lower total and the most atherogenic LDL-cholesterol by up to 50-55% at maximum dosage. The beneficial effect of statins on plasma lipoproteins is reflected by a significant reduction in lesions progression and major cardiovascular events (by 25-40%) as demonstrated by many prospective primary and secondary prevention trials. The fibrates, another class of LLA that primarily decrease plasma triglycerides and increase the anti-atherogenic HDL-cholesterol, have also shown in angiographic and event trials a 25% reduction of major cardiovascular end points. More recently, the use of a statin to decrease LDL-C in combination with nicotinic acid to increase HDL-C in secondary prevention gave an even more spectacular 90% reduction of recurrent cardiovascular events after three years of treatment.

If the question of the benefit of treatment with LLA on cardiovascular disease is no more debatable, many issues regarding the use of LLA in clinical practice are still unresolved. For instance, in vitro and animal studies have documented various pleiotropic beneficial effects of statins on coagulation, fibrinolysis, oxidation and inflammation. Since these pathophysiological mechanisms are fundamental to atherosclerosis progression, it will be of great interest to assess whether the beneficial effect of statins is only related to lipid lowering or it is also mediated by other mechanisms. In patients with low lipid levels who are at risk of CHD because of intravascular inflammation or oxidation it remains unknown whether they would benefit from statins. Another fundamental question regards the optimal level of plasma cholesterol. It is still questioned whether the “lower is better” and whether there is a threshold effect to the lipid lowering benefit. Obviously many combination therapy trials will have to be conducted to resolve this issue.

Concerning the long term safety of LLA and recognition of adverse effects little is known. The longer these agents are used, the more we learn about subtle presentation of side effects. The neuromuscular symptoms associated with statin use or the fibrate induced fatigue are a few examples. These unrecognized side effects are also certainly involved in the problem of poor compliance to LLA. It is well documented, particularly in the elderly population, that after one year of treatment only 25% of the patients are still taking their statin. Even more worrisome is the possibility that a patient may suffer an unrecognized side effect that limits their quality of life. The issue of safety is crucial to permit the use of LLA in the paediatric population. Since statins provide effective prevention of atherosclerosis and CHD to patients with genetic diseases such as familial hypercholesterolemia in which cholesterol deposition in arteries starts in childhood, it is mandatory to establish the safety of statins at younger ages.

Some forms of morbid dyslipidemias such as hyperchylomicronemia and secondary pancreatitis are still untreated by drugs. Fortunately microsomal triglyceride transfer protein (MTP) inhibitors are in development and will need to be tested in these severe hypertriglyceridemic conditions. Other agents targeting enzymes and receptors of lipid metabolism such as cholesterol transfer protein (CETP) inhibitors, acyl transfer protein inhibitors, bile acid transporter inhibitors, LDL antioxidants etc. will also increase our arsenal of LLA in the next few years. These drugs will certainly open the way to combination therapies but may be also be the cause of lipid lowering drug interactions.

II. PHASE II STUDIES FOR REGISTRATION OF NEW DRUGS

II.1. Outline of a typical development plan

During this phase the lipid lowering effect of the new drug is compared to placebo in a population of moderately hyperlipidemic patients over a relatively short period of time (12-24 weeks). These controlled trials are initiated once earlier studies in normal volunteers have assessed the pharmacokinetics, optimal and
maximally tolerated dosages, dose schedule and interactions with food and other drugs. The Phase II trials are designed to evaluate the efficacy and short-term safety of the new agent. The assessment of efficacy of lipid lowering drugs is not free of ambiguity. LDL-C is an accepted surrogate for coronary heart disease, the ultimate aim of lipid lowering, and thus evidence that a drug lowers LDL-C by at least 15% is adequate for registration. However, for HDL-C raising and TG lowering drugs, not proven surrogates for CHD, it is likely that in addition to showing a benefit on lipoproteins, the demonstration of a benefit on clinical CHD outcomes will also be needed.

A multicenter, randomized, parallel group design with 2 to 4 groups receiving several drug doses, and one placebo group is generally used. Regulatory agencies require a minimum of two drug dosages and 12 weeks of active treatment for approval. A minimum of two dosages is necessary for observation of a dose versus effect relationship. The number of patients randomized is dependent upon the magnitude of the lipid lowering effect and the drop out rate expected. Patients with secondary dyslipidemia and a history of cardiovascular disease (secondary prevention) are excluded since the risk of placebo treatment for 3-4 months in this last group, is considered unethical. Because treatment with LLA is always additive to diet and lifestyle modifications, a pre-randomization run-in period of 4-8 weeks on diet and placebo is mandatory. Patients included in these short-term trials are usually allowed to enter subsequent long-term follow-up trials.

II.2. Short-term studies

Monotherapy trial for treatment of primary dyslipidemia

II.2.A. Objectives
   a. To determine the safety and efficacy as monotherapy for primary hypercholesterolemia.

II.2.B. Primary endpoints
   a. The percent change of LDL-cholesterol from baseline (randomization visit) to the end of the study.
   b. The incidence and prevalence of adverse events and clinical safety laboratory parameters abnormalities.

II.2.C. Secondary endpoints
   a. Percentage change in lipoprotein cholesterol and triglycerides (total, LDL-C, HDL-C, VLDL-C).
   b. Percentage change in apolipoproteins (apoA1, apoB, LDL apoB and Lp(a)).
   c. Percentage of patients meeting the NCEP (National Cholesterol Education Program) guidelines target lipid values at the end of the study.
   d. Percentage change in fat-soluble vitamin levels (vitamin E, A, D, β-carotene) and INR as a functional measure of vitamin K status (for drugs affecting lipid absorption).
   e. Quality of life assessed by the SF-12 Health Survey.

II.2.D. Study design
A multicenter, randomized, double-blind, parallel group, placebo controlled design is generally used. Study begins with a screening phase using patient’s file, questionnaire, physical examination and blood tests to identify those meeting the clinical and lipids inclusion/exclusion criteria. Potential candidates then enter a 4-8 weeks lead-in, stabilization period, during which they will receive single-blind placebo medication, lifestyle and diet counseling according to NCEP guidelines. A minimum of six weeks without LLA is needed before randomization. The lead-in period is also used to monitor drug compliance. Those not taking 75-80% of their medication are usually excluded. At this point patients fulfilling the inclusion/exclusion criteria are randomized to placebo or one of two active treatment arms (with one or up to four different doses of study medication). Sometimes a short titration phase may be used within the active treatment arms. During the treatment period lasting generally 12 to 24 weeks, the dosage remains stable until the end of the study. No other lipid lowering or lipid affecting agents are allowed during the
study, and an effort is made to keep the concomitant medications stable. At the end of the study, each participant is usually invited to take part in a long term extension trial with a predetermined randomization to an active drug treatment regimen.

II.2.E. Planned sample
Assuming a dropout rate of 15 to 25% over a 24 weeks period, approximately 150 patients per group are needed to detect, with 90% power, a significant difference of 15% ($p \leq 0.05$) between each treatment arm and placebo.

II.2.F. Study population
Adults (over 18 years) with primary hypercholesterolemia.

II.2.G. Specific inclusion criteria
a. Adults over 18 years of age at screening.
b. Females must not be pregnant or lactating. Females of childbearing potential, must use a medically acceptable form of contraception at least four weeks before and until four weeks after the end of the study.
c. Documented history of hypercholesterolemia with LDL-cholesterol (a mean of the 2 values obtained at the screening and randomization visits) $\geq 3.4$ mmol/L and $\leq 6.2$ mmol/L and TG $\leq 3.5$ mmol/L.
d. Have the ability to comply with the NCEP Diet.

II.2.H. Specific exclusion criteria
a. History of atherosclerotic vascular disease including CABG (Coronary Artery Bypass Graft), PTCR (Percutaneous Transluminal Coronary Revascularisation), clinical or symptomatic angina pectoris, myocardial infarction, stroke, or peripheral vascular disease.
b. Uncontrolled primary hypothyroidism (as defined by TSH $> 1.5$ times the upper limit of normal at the screening visit), nephritic syndrome, or other possible causes of secondary dyslipidemia.
c. HemoglobinA1C $> 8.0\%$.
d. History of cancer in the past five years (excluding basal cell carcinoma).
e. Uncontrolled hypertension as defined by a systolic BP $\geq 160$ mmHg or diastolic BP $\geq 100$ mmHg.
f. Chronic renal failure or serum creatinine $> 1.7$ times the upper limit of normal at the screening visit.
g. Unexplained serum CK $> 3$ times the upper limit of normal at the screening visit.
h. Active hepatitis or cholestasis or ALT $> 2$ times the upper limit of normal at the screening visit.
i. History of drug or alcohol abuse within the last year (more than 21 alcoholic beverages/week).
j. Chronic diarrhoea or malabsorption.
k. Subjects taking one of the following medications and unable to maintain a stable dose at least four weeks prior to the screening visit and for the duration of the study: tamoxifen, raloxifene, estrogen and or progestins, thiazide diuretics, isotretinoin, β-blockers, thyroid hormones, androgens, fiber supplements, protease inhibitors.
l. Subjects currently taking and unable to discontinue prior within eight weeks prior to randomization and throughout the study: Any lipid lowering agent, cyclosporine, orlistat, systemic corticosteroid, alpha-glucosidase inhibitors,
m. Drug compliance lower than 80% of the expected tablet count during the last two weeks of the lead-in period.
n. Have received any investigational drug within four weeks prior to the screening visit.

II.2.I. Tools for assessing primary endpoints
Blood tests.
II.2.J. Specific criteria for early withdrawal and discontinuation
The criteria for discontinuation from the study are pre-specified:
   a. Failure to meet randomization criteria;
   b. Protocol non-compliance;
   c. Adverse events;
   d. Investigator judgment;
   e. Patient withdraws consent;
   f. Pregnancy;
   g. ALT $\geq$ 3 times the upper limit of normal;
   h. CK $\geq$ 10 times the upper limit of normal or CK $\geq$ 5 times the upper limit of normal with clinical signs of myopathy; on two consecutive occasions at least one week apart.

II.2.K. Data analysis method
Baseline homogeneity of the variables is examined by a one way ANOVA. The analysis of efficacy is done according to the intention-to-treat principle. All statistical tests are two sided and p values $\leq 0.05$ are considered statistically significant. The only multiplicity adjustments are the Bonferroni adjustments for the primary hypothesis tests, one for each of the active treatment arm vs. the placebo.

II.3. Long-term studies
Following completion of the short term, double blind, placebo controlled, efficacy and safety study, subjects are often offered entry into an open-label, long term (usually 12 months), extension study. The purpose of the trial being to gather data on long-term safety and tolerability of the study drug. In monotherapy long-term trials, during the first few weeks, the lipid lowering drug is usually titrated up (when it is possible), for subjects not attaining at the end of the short-term trial, the LDL-cholesterol goal dictated by NCEP guidelines. In combination trials, it may be the combined agent that is titrated up. After the titration period, usually patients are followed less frequently (every 3 months). Still, if necessary to reach therapeutic goal, other lipid lowering agents may be added during long-term follow-up. To demonstrate safety, the ICH guidelines generally ask that 600 patients be treated for 6 months with the new agent and 100 patients studied for one year. These extension studies offer also the possibility to test the persistence of the lipid lowering effect.

III. PHASE III STUDIES FOR REGISTRATION OF NEW DRUGS

III.1. Outline of a typical development plan
The phase III development plan of a lipid lowering drug must include a double-blind, randomised, placebo-controlled parallel group study which replicate the short-term phase II study, but generally uses only the optimal dosage and recruits a larger number of patients. The study is performed in duplicate usually in different populations; one in North America and one in Europe for example. The U.S. Food and Drug Administration (FDA) requires that phase III studies need to recruit a significant number of women and non-Caucasian participants. The main objective of phase III studies is to test the reproducibility and the expandability of the phase II results to the general population.

The number and the nature of the other trials in the development plan will be dictated by the intended use and indication of the drug. For instance, if a strong and well tolerated hypocholesterolemic agent with a new mechanism of action is intended to be used in replacement of statins to decrease LDL-cholesterol lowering, the phase III plan will include:
   a. Trials comparing the most powerful statins on the market with the new agent, in moderately and severely hypercholesterolemic subjects.
b. The same comparison in special populations such as patients with Heterozygous Familial Hypercholesterolemia (FH) adult or children, Homozygous FH, renal failure, sitosterolemia or aortic stenosis.

c. Combination trials with fibrates in patients with mixed dyslipidemia.

d. Long-term outcome trials (recurrent myocardial infarction or cardiovascular death) in secondary prevention.

e. Trials measuring the benefit of the drug on surrogate markers of atherosclerosis such as endothelial dysfunction, intima-media thickness or coronary plaques evaluated by intravascular ultrasound (IVUS) or angiography.

Before testing the new drug in children, elderly people or patients with renal failure it is necessary to demonstrate the pharmacokinetic and pharmacodynamic equivalence in a few subjects (8-10). A trial performed in a paediatric population with Familial Hypercholesterolemia is outlined below as an example of a phase III study.

III.2. Short-term monotherapy studies

Monotherapy trial in children with Heterozygous Familial Hypercholesterolemia

III.2.A. Objectives
To evaluate the efficacy and safety as a monotherapy agent in adolescents with heterozygous familial hypercholesterolemia (HeFH).

III.2.B. Primary endpoints
a. Percentage of change of LDL-cholesterol from baseline to six months.
b. Incidence of adverse events, ECG and clinical laboratory safety parameters.

III.2.C. Secondary endpoints
a. Specific apolipoproteins, lipoprotein cholesterol and triglycerides.
b. Serum steroid hormone levels
c. Serum fat-soluble vitamin levels
d. Serum vitamin B-12 and serum/red blood cell folate levels
e. INR as a functional measure of vitamin K status
f. Linear growth

III.2.D. Study design
A multi-center, double-blind, placebo-controlled, parallel group, randomized study to evaluate the safety and efficacy in monotherapy. The study consists of a five to eight week drug washout, diet stabilization lead-in period, and a twenty-six (26) weeks double-blind treatment period. At the end of the lead-in period, patients meeting the randomization criteria will be randomized to one of the two arms: Active drug at the optimal dose or placebo. An unbalanced randomization can be used such that the ratio of patients on active drug/placebo will be 2/1.

III.2.E. Planned sample
Sample size will be similar to what has been described for phase II monotherapy trials. But since the study includes adolescents, it is wise to recruit a larger number of patients to counteract a significant Lost to follow-up. With the same purpose, the number of visits should be kept minimal.

III.2.F. Study population
Male and female adolescents between 10 and 17 years of age with HeFH.
III.2.G. Specific inclusion criteria
   a. HeFH documented by LDL-cholesterol levels consistent with FH (LDL-C > 4.2 mmol/L) or documentation of LDL-cholesterol receptor gene mutation AND a positive family history of atherosclerosis at or before 50 years of age in males, and before 60 years of age in females, OR documented family history of hyperlipidemia with LDL cholesterol levels above 95th percentile for age and sex before treatment.
   b. The participant and the parent/legal guardian must provide informed consent prior to the subject undergoing any study-specific screening procedures.

III.2.H. Specific exclusion criteria
   a. As described in the monotherapy phase II trial for treatment of primary dyslipidemia.
   b. History of homozygous familial hypercholesterolemia.

III.2.I. Tools for assessing primary endpoints
   Blood tests.

III.2.J. Specific criteria for early withdrawal and discontinuation
   As described in the monotherapy phase II trial for treatment of primary dyslipidemia.

III.2.K. Data analysis method
   As described in the monotherapy phase II trial for treatment of primary dyslipidemia.

Long-term monotherapy studies
   As described in the monotherapy phase II trial for treatment of primary dyslipidemia.

IV. PHASE III STUDIES FOR REGISTRATION OF NEW DRUGS: ADJUNCTIVE THERAPY

IV.1. Outline of a typical development plan
   At the present time, the vast majority of patients treated for dyslipidemia receive only one lipid agent (monotherapy). There are numerous reasons for that: First, the NCEP has designated LDL-C reduction as the primary target of treatment to prevent atherosclerosis and statins used in monotherapy can normalize LDL-C for most of the patients. Second, most of the long-term event trials that have proven the benefit of lipid lowering to prevent CVD have been done with monotherapy. Third, combinations of certain lipid lowering agents (i.e. gemfibrozil+ statin) have been proven hazardous in the past, with cases of rhabdomyolysis and acute renal failure. Fourth, in patients with mixed dyslipidemia (elevated LDL-C and TG), the hypertriglyceridemia is often secondary to overweight, a poor diet, lack of exercise or alcohol intake, and can be corrected by lifestyle modifications. Fifth, physicians are reluctant to use two different drugs in asymptomatic patients to reduce, at significant costs, a single risk factor of CVD, particularly in primary prevention.

   Nevertheless, some clinical situations require combinations of lipid lowering agents to effectively treat morbid forms of dyslipidemias. A combination of a full dose statin and ezetimibe is often necessary to normalize LDL-C in HeFH patients. A statin associated with nicotinic acid can effectively take care of high LDL-C and low HDL-C, a form of dyslipidemia often encountered in survivors of myocardial infarction. High risk patients with severe mixed dyslipidemia may need a combination of a fibrate and a statin for optimal therapy.
New drugs in development may have relatively modest cholesterol lowering effect but a complementary anti-atherosclerotic mode of action on oxidation, inflammation or thrombosis. The phase III development design for such agents intended to be used as adjunctive therapy must minimally include:

a. Two large scale, short-term, parallel group, trials in different populations comparing the effect of the optimal dose of the new drug used in combination with a statin vs the statin alone on lipid levels and surrogate markers of atherosclerosis, inflammation and oxidation.

b. Another short-term trial using the new drug in combination with a fibrate compared with the fibrate alone in patients with mixed dyslipidemia.

c. A long-term trial comparing in secondary prevention, the same treatment arms over cardiovascular endpoints.

d. A long-term trial testing the effect of the combination with a statin on carotid atherosclerosis progression evaluated by measurement of intima-media thickness (the only method approved by FDA).

Some studies on LLA intended to be utilised in combination therapy have used a cross-over design where each patient is treated successively with each drug and then with the combination. In such study design each patient serves as his own control, and each of the three treatment periods has to be long enough (2-3 months) to prevent a carry-over effect of the previous treatment.

**IV.2. Short-term adjunctive therapy studies**

**Combination-therapy trial in subjects with primary hypercholesterolemia at high-risk of cardiovascular disease not controlled by a starting dose of a statin alone**

**IV.2.A. Objectives**

To evaluate the efficacy and the safety of the investigational drug administered in combination with a statin in subjects with primary hypercholesterolemia and at high risk of cardiovascular disease, who do not reach the target value of LDL-C (2.5 mmol/L) on starting dose of the statin alone.

**IV.2.B. Primary endpoint**

a. The proportion of subjects achieving target LDL-C (according to NCEP ATP-III guidelines) after 12 weeks of treatment (the aim is to show that a greater number of patients will reach the LDL-C target by combining the investigational agent to the statin rather than by increasing the dosage of the statin).

**IV.2.C. Secondary endpoints**

b. The proportion of subjects achieving target LDL-C at weeks 2, 4 and 8.


d. Change in quality of life, evaluated by the SF-36 questionnaire, after 12 weeks of treatment.

e. Adverse events and laboratory abnormalities occurring during the active treatment phase.

**IV.2.D. Study design**

A multi-center, randomized, double-blind, parallel group study.

Subjects with primary hypercholesterolemia and at high risk of cardiovascular event because of CHD history or diabetes mellitus or presence of other risk factors (absolute risk greater than 20% at 10 years according to Framingham tables) will be recruited. After a four weeks wash-out period of any lipid lowering agent (eight weeks for fibrates and one year for probucol), candidates will receive a starting dose of statin for another four weeks (open-label). At the end of the open-label period, they will be randomized to continue on statin therapy and placebo or to the combination of statin and the investigational drug at a
fixed dose for 12 weeks. In each group, the statin dosage is doubled after 4 and 8 weeks of treatment if LDL-C remains greater than 2.5 mmol/L. Diet therapy will be followed for the duration of the study.

**IV.2.E. Planned sample**
To show a 15% difference in the percentage of subjects achieving the target LDL-C level, with a power of 90% and a significance level of 0.05; a total sample of approximately 500 participants (250 per treatment arm) is needed.

**IV.2.F. Study population**
Men and women with primary hypercholesterolemia as well as with cardiovascular disease or well controlled diabetes mellitus or a cardiovascular risk greater than 20% at 10 years according to Framingham tables (NCEP ATP-III)

**IV.2.G. Specific inclusion criteria**
   a. Age 30 to 70 years.
   b. Primary hypercholesterolemia with an LDL-C ≥ 2.5 mmol/L and TG ≤ 3.5 mmol/L after the four weeks run-in period of open label statin therapy.
   c. Documented history of cardiovascular event (acute coronary syndrome; coronary, cerebral or peripheral revascularization; stroke or transient ischemic attack).
   AND/OR
      Well controlled and stable diabetes mellitus (HbA1C ≤ 7.0% at baseline and no change in treatment later than 3 months before screening).
   AND/OR
      A risk of CHD greater than 20% at 10 years, according to the modified Framingham tables of the NCEP ATP-III.

**IV.2.H. Specific exclusion criteria**
   a. Medical condition likely to limit life span to less than one year
   b. Women of child bearing potential who do not agree to practice an effective method of birth control until one month following study completion
   c. Lactating women
   d. Postmenopausal women on hormone replacement therapy
   e. Patients with homozygous familial hypercholesterolemia
   f. Uncontrolled endocrine or metabolic disease known to influence serum lipids or lipoproteins (thyroid hormone substitution must be stable for at least 3 months prior to screening)
   g. Chronic renal insufficiency or nephrotic syndrome
   h. Active or chronic hepatobiliary or hepatic disease
   i. Treatment with agents with known interactions with the statin or the investigational drug used in the study
   j. Any other lipid-altering agents (drug or food supplement) administered during the study

**IV.2.I. Tools for assessing primary endpoints**
Blood tests and questionnaire about quality of life (SF-36 Acute, Chapter 25, Appendix B).

**IV.2.J. Specific criteria for early withdrawal and discontinuation**
Threatening of health or well-being of a participant by their continuation in the study, in the opinion of the investigator.

**IV.2.K. Data analysis method**
The analysis of the primary efficacy variable is based on the intention-to-treat and performed using chi-square test. Analysis at each time point (4, 8 and 12 weeks) is provided.
IV.3. Long-term adjunctive therapy studies

Combination-therapy trial in subjects with mixed dyslipidemia

IV.3.A. Objectives
To demonstrate that the fixed combination of a new triglyceride lowering agent (TLA) with a cholesterol lowering agent (CLA), titrated to the NCEP ATP-III guidelines, can slow the progression of atherosclerosis when compared to the cholesterol lowering agent alone, in subjects with mixed dyslipidemia after 24 months of treatment.

IV.3.B. Primary endpoint
a. Annualized rate of change in intima-media thickness (IMT).

IV.3.C. Secondary endpoints:
b. Annualized rate of change in IMT of each anatomical site (internal carotid, bifurcation and common carotid).
c. Change in lipid parameters and biochemical markers of vascular inflammation and insulin resistance.
d. Cardiovascular events (combination of coronary death, non-fatal myocardial infarct, stroke and revascularization procedures).
e. Adverse events and abnormalities of clinical laboratory parameters.

IV.3.D. Study design
A multi-center, double-blind, randomized, parallel group, carotid ultrasound study of the fixed combination compared to the cholesterol lowering agent alone administered to subjects with mixed dyslipidemia. Potential subjects will undergo a 4-week lead-in period when they will receive diet and lifestyle counselling only (no LLA), as described in the NCEP ATP-III clinical guidelines or equivalent to establish screening lipid levels. This is especially important in patients with mixed dyslipidemia since diet and lifestyle has great impact on triglyceridemia. At the end of the lead-in period, eligibility for entry into the study is determined. All eligible subjects will enter a run-in period of treatment with the cholesterol lowering agent alone. The dosage will be titrated-up every four weeks to a target LDL-C level as per NCEP ATP-III guidelines. Once the target LDL-C is reached, subjects will then be randomized to continue on the CLA alone or to the combination with the TLA at fixed dose. Carotid ultrasonography will be performed at baseline and every six months for 24 months of treatment. Clinical safety and/or lipid efficacy assessments will be performed at each visit.

IV.3.E. Planned sample
It is estimated from similar studies that the standard deviation of the annualized rate of IMT change is 0.06 mm/yr. With a sample size of 340 subjects in each treatment group and a two-sided alpha level of 0.05, the study will have 90% power to detect a difference of 0.015 mm/yr. With a dropout rate of 20%, 850 subjects will be randomized.

IV.3.F. Study Population
Men and women with mixed dyslipidemia persisting after diet and lifestyle modifications. Patients with dysbetalipoproteinemia (Type III) should be excluded because they are particularly responsive to TLA.

IV.3.G. Specific inclusion criteria
a. Age 30 to 65 years
b. Mixed dyslipidemia defined by LDL-C ≥ 4.2 mmol/L and ≤ 6.2 mmol/L AND TG ≥ 2.3 mmol/L and ≤ 6.5 mmol/L.
IV.3.H. Specific exclusion criteria

a. Women who are pregnant or lactating.
b. Patients with dysbetalipoproteinemia (Type III)
c. Patients with symptomatic cardiac or cerebrovascular disease (within the last 3 months prior to the screening visit)
d. Patients who do not have a satisfactory baseline carotid ultrasound.
e. Patients requiring specific lipid lowering therapy different from the CLA used in the study.
f. Subjects taking any drugs known to interact with the medications used in the study.
g. Subjects with uncontrolled hypertension (>140/90 mmHg), diabetes or hypothyroidism.
h. Subjects with history of acute pancreatitis, malignancy, hepatobiliary disease (active), nephrotic syndrome, chronic renal failure or malabsorption.
i. Subjects with alcohol consumption of more than 14 alcoholic beverages per week.
j. Subjects with severe obesity or body mass index > 35 kg/m2.
k. Subjects with poor compliance during the run-in period (< 80%) as assessed by tablet count.

IV.3.I. Tool for assessing primary endpoints
Quantitative ultrasonography.

IV.3.J. Specific criteria for early withdrawal and discontinuation
Finding of severe atherosclerotic carotid stenosis requiring surgery or adverse event requiring discontinuation of treatment.

IV.3.K. Data analysis method
The analysis of safety and efficacy data is done on the intention-to-treat population. But to be included in the efficacy analysis a participant needs to have at least both a valid baseline and a valid follow-up carotid IMT evaluation performed at least one year post baseline. The primary endpoint comparison between the two treatment groups needs to include analysis of time by treatment interaction.

V. OTHER STUDIES (SPECIAL INDICATIONS)

V.1. Hyperlipidemia in the elderly
Since the benefit of treating hyperlipidemia to prevent cardiovascular events in elderly men and women (70-80 years) at high cardiovascular risk has been demonstrated in large scale randomized trials; the population involved in new drug trials should include a significant proportion of older patients. The number of drugs susceptible to be prescribed unilaterally to older people for secondary prevention of CVD is growing every year (aspirin, ACE inhibitors, beta-blockers, clopidogrel, statins, etc.). As a consequence, pharmacokinetic and interaction studies should be performed in elderly subjects. New drug phase III trials should not be restrictive in terms of concomitant medication in order to represent a “real life” situation and obtain valuable information on the safety of the new agent in this segment of the population. It may also be judicious in the future to include systematically in every long term trial involving subjects older than 70 years an evaluation of cognitive functions in order to detect subtle beneficial or deleterious effects.

V.2. Hyperlipidemia in dialysis and renal transplant patients
Total cholesterol level is inversely associated with survival in dialysis patients, a group at high risk of cardiovascular disease. The nephrotic syndrome is associated with high plasma levels of total and LDL-C, thus can promote atherogenesis and the risk of CHD. Chronic renal failure is characterized by hypertriglyceridemia, and decreased levels of HDL-C in approximately 30% of patients. Patients with kidney transplantation exhibit variable patterns of hyperlipidemia; some patients are purely hypertriglyceridemic while others present with hypercholesterolemia or mixed dyslipidemia. Unfortunately, the risk of myotoxicity and of interaction between cyclosporine and many of the currently available lipid lowering drugs (statins or fibrates), and the lack of large scale trials, has hampered their
utilization in this subgroup of patients. Thus, new drug development projects should include phase III short and long term studies to examine the efficacy and safety to reduce hyperlipidemia and CVD in patients with renal disease.

VI. EXAMPLES OF LANDMARK WELL DESIGNED TRIALS

Monotherapy, short-term trials

Monotherapy, long-term trials

Adjunctive therapy trials

Clinical events trials
Chapter 13. Anti-atherosclerotic Drugs

Jean-Claude Tardif, M.D. ¹

Therese M. Heinonen, D.V.M. ²

Marie-Claude Guertin, Ph.D. ²

¹ Research Center
Montreal Heart Institute
5000 Belanger Street
Montreal, Quebec H1T 1C8
CANADA
Tel 514-376-3330 x 3612
Fax: 514-593-2500
Email jean-claude.tardif@icm-mhi.org

² Research Center
Montreal Heart Institute
5000 Belanger Street
Montreal, Quebec H1T 1C8
CANADA
1. INTRODUCTORY REMARKS

At this moment, atherosclerotic disease places a great burden on society. Not only is atherosclerosis currently one of the leading causes of death around the world, but it also has a major impact on an individual’s quality of life as a result of chronic pain, activity restriction, unemployment and disability. In 2002 in the United States alone, it was estimated that over 62 million individuals had one or more type of cardiovascular disease and that the direct and indirect cost of atherosclerotic coronary artery disease (CAD) was over 110 billion dollars.

A variety of risk factors for the development of atherosclerosis have been identified. These risk factors include smoking, high blood pressure, high cholesterol, diabetes, obesity, male gender, genetic predisposition, and physical inactivity. Lifestyle modifications and pharmacological therapy, including cholesterol-lowering medication, have become the standard of care provided to patients with coronary artery disease. While these interventions have been effective, they have failed to keep pace with the expanding “at risk” population. In particular, the increased proportion of elderly individuals and the rise in prevalence of risk factors such as metabolic syndrome, diabetes, and obesity are cause for concern. It is important to recognize that cardiovascular disease not only affects the elderly but is also a leading cause of premature death in individuals under the age of 75. Furthermore, hospitalization rates by age groups indicate that acute myocardial infarction (MI) and ischemic heart disease become important diseases by the fourth and fifth decades of life. As a result of this trend, there has been an increase in the number of deaths due to cardiovascular disease worldwide with the expectation that this trend will continue for the next fifteen years.

A complex series of events leads to the formation of atherosclerotic plaques in human arteries. Atherosclerosis is now understood to be a systemic chronic inflammatory disease characterized by the excess accumulation of lipid-laden macrophages (foam cells) within the arterial wall. The steps involved in the development of atherosclerotic lesions include endothelial expression of adhesion molecules, release of cytokines and chemokines, involvement of reactive oxygen species, macrophage accumulation in the arterial wall, and incorporation of oxidized LDL. The resulting atheroma is comprised of cholesterol, inflammatory cells, and matrix. Atherosclerotic lesions are heterogeneous and the clinical manifestation of atherosclerosis is dependent on the lesion composition and affected vascular bed, as well as other factors. When the atherosclerotic plaque ruptures the clinical presentations include MI, unstable angina, stroke, transient ischemic attack (TIA), and cardiac death.

The complex pathophysiology of atherosclerotic disease highlights the fact that many processes contribute to lesion development, and suggests that cholesterol modification is not the only mechanism by which the condition may be positively affected. It is well accepted that high serum levels of cholesterol, and particularly low-density lipoprotein cholesterol (LDL-C), increase the risk for CAD. The use of HMG-CoA reductase inhibitors (statins) became widespread as data from numerous clinical trials supported their safe and efficacious use in a broad population base. With a reduction in combined cardiovascular morbidity and mortality of 30 – 35% resulting from their use, statin drugs have had a substantial positive benefit on health care. However, the majority of clinical events are not prevented by statins and the persistently high level of cardiovascular disease is far from satisfying. Attention has therefore been focused on non-lipid risk factors of CAD including some of the key systems involved in the formation and progression of atherosclerotic plaque. As a result of the ubiquitous and chronic nature of atherosclerosis, as well as the persistently high incidence of cardiovascular morbidity and mortality, research efforts in this area remain vital. In recent years these efforts have changed our understanding of atherosclerosis significantly. A process which was once thought of as a static accumulation of fat in the arteries is now realized to be a very dynamic process in which numerous environmental, genetic, and individual factors contribute. Research which at one time focused almost exclusively on cholesterol metabolism has now branched out to provide further insight into the role of
other processes such as inflammatory contributors. A number of chemokines and cytokines involved in the inflammatory process have now been identified, and potential markers of disease progression and regression are under evaluation. Additionally, our ability to understand the process of atherosclerosis and the causative factors has gained sophistication with the experimental and clinical employment of various high-technology tools. These tools include the use of genomics, proteomics, the identification of soluble plasma markers, as well as imaging technologies such as quantitative coronary angiography, intravascular coronary and peripheral ultrasound, B-mode carotid ultrasound, positron emission tomography, computed tomography, and magnetic resonance imaging.

The development of atherosclerotic plaque changes both vascular structure and function. The endothelial lining of the blood vessels is important in the regulation of vascular tone and the accommodation toward greater blood supply in circumstances of greater demand. Modification of endothelial function is one of the earliest clinically demonstrated changes within the vasculature during the course of atherosclerotic progression. Subsequent lesion progression with the accumulation of cholesterol and matrix leads to clinically detectable changes in vascular anatomy and further functional changes. The vascular wall becomes thickened with accumulating debris, and the lumen eventually narrows despite remodelling changes to accommodate expanding plaque volume. The convergence of these factors, along with the inflammatory components, results in supply/demand mismatch as well as plaque instability and rupture—all accumulating in clinical disease.

While progress has been made in the field of atherosclerosis, there remain many challenges in developing an agent which will work by a non-lipid anti-atherosclerotic mechanism. There are currently no regulatory guidelines for the development of such drugs and no therapies approved for the treatment of atherosclerotic coronary disease that work by non-lipid mechanisms. Additionally, despite the recent generation of data identifying potential biomarkers such as C-reactive protein, clear and independent links between such markers and coronary artery disease still need to be strengthened in order to maximize their utility in drug development programs. Other developmental hurdles include the chronic nature of CAD, trial duration, ethical considerations, concomitant medications, and plaque heterogeneity. Since atherosclerotic plaques develop slowly over decades, in most instances, clinically meaningful changes using even the most sensitive methods have required relatively long-term therapy of 12-18 months. Due to the effectiveness of current therapy, including statins, at decreasing cardiovascular risk in a wide patient population, true placebo-controlled trials in patients with cardiovascular risk have become unethical. Also due to the demonstrated effectiveness of current therapy, any potential new therapy is expected to be evaluated as an adjunct to standard care. Standard care may involve several medications for lipid-lowering, as well as hypertension and glucose control. Care must therefore be given to potential interactions between new agents and these concomitant medications. While the development of new anti-atherosclerotic agents that work by non-lipid mechanisms will continue to present a developmental challenge, the new tools available, and the diligence of those undertaking this challenge is expected to make significant inroads in our ability to successfully identify the next wave of effective therapy.

II. PHASE II STUDIES FOR REGISTRATION OF NEW ANTI-ATHEROSCLEROTIC DRUGS

II.1. Outline of a typical development plan

Although atherosclerosis is a systemic condition, which leads to stroke and other serious peripheral circulatory disorders, the development of anti-atherosclerotic agents typically focuses on coronary disease indications. Standard exploratory studies conducted during this stage of development include those designed to assess the maximally tolerated dosages over a short to moderate treatment durations, pharmacokinetics and drug interactions. The length of treatment should be sufficient to obtain the safety and efficacy information needed to determine the feasibility of the long-term evaluations needed in Phase III. Additionally, since these agents work by a non-lipid mechanism, it is important for the early clinical
trials to explore the utility of other potential biomarkers. Two types of biomarkers may be included in these trials. The first type of biomarker is one which will provide pharmacodynamic information on the biochemical system targeted. For example, an agent which activates enzyme X should be evaluated at this stage for the degree of activation conferred across an acceptable dose range, if possible. The second type of biomarker is one which will provide a reasonable indication of clinical efficacy (i.e. vascular improvement). The selection of potential biomarkers for evaluation should be based on a good understanding of the agent’s mechanism of action, as well as the pathophysiology of atherosclerotic disease. The appropriate set of biomarkers for each evaluated agent may be quite different depending on the mechanism and anticipated response. Therefore the development plan for each agent may also be unique. Additionally, since lipid-lowering therapy has clearly demonstrated effectiveness and safety, it is important to assess at an early stage the influence of the study drug on current therapy through the use of well designed drug interaction studies. The results of these carefully designed early studies are critical for appropriate design of pivotal studies, which may be significant in terms of size, duration, and resource needs.

There are several different approaches employed in Phase II to obtain clinical efficacy. One approach is to evaluate a population with a clinical indication for cardiac catheterization in order to enable the use of sensitive but invasive imaging tools such as intravascular ultrasound (IVUS) to detect drug-induced changes. Such an approach has the advantages of directly assessing the coronary circulation, minimizing sample size and detecting small but potentially important changes in plaque burden. However, the obvious disadvantages are the invasiveness of the assessment and the limitations regarding patient selection. Another approach is to conduct initial evaluations in individuals with stable moderate atherosclerotic disease using non-invasive tools such as B-mode ultrasound in peripheral vessels (carotid and brachial). Such an approach has the advantage of non-invasive assessments in a widely available patient population, but relies on correlations between peripheral changes and coronary disease. Additionally, this approach may require a larger sample size to detect drug effect changes. Often, both of these approaches are included in the Phase II development plan.

The typical pivotal trial will utilize a double-blind, multi-center, placebo-controlled, randomized, adjunctive therapy, parallel group design. At least 2 or 3 dose levels should be explored, preferably within the same trial. Early proof-of-concept monotherapy studies or single-center studies may be included as part of the phase II development program. Patients included in short-term studies should be allowed to enter long-term open-label follow-up for collection of additional safety information.

II.2. Short-term studies

II.2.i. Adjunctive therapy trial in patients with clinical indication for cardiac catheterization - IVUS

II.2.i.A. Objectives
To evaluate short-term efficacy and tolerability during adjunctive therapy use

II.2.i.B. Primary endpoints
a. The percent change in plaque volume (follow-up – baseline)/baseline x 100) measured by 3D IVUS.

II.2.i.C. Secondary endpoints
a. Absolute change in plaque volume on three dimensional (3-D) IVUS
b. Change in percent plaque volume on 3-D IVUS
c. Changes in plaque volume in anatomically comparable 5-mm segments centered on the sites with lowest and highest plaque burden at baseline by 3-D IVUS
d. Changes in plaque characterization indices assessed by IVUS
122

e. Coronary score assessed by quantitative coronary angiography (QCA) defined as the per-patient mean of the minimal lumen diameter for all lesions measured.
f. Incidence of adverse events

II.2.i.D. Exploratory endpoints
Changes from baseline in markers of biochemical target activation and markers of atherosclerosis pathophysiology (ie. inflammatory markers).

II.2.i.E. Study design
A multicenter, randomized, placebo-controlled, parallel-group design is typically used. Inclusion criteria usually require patients to be at least 18 years of age and with a need for diagnostic coronary angiography for clinical indication (with or without percutaneous coronary intervention). Patients are usually required to have stable background treatment following standard care practice. Patients are screened and the initial IVUS examination is performed in a target coronary artery which must not have undergone previous percutaneous coronary intervention (PCI) nor be a candidate for intervention at the time of the baseline catheterization. The baseline phase should be of sufficient duration to allow for scheduling of catheterization procedures within an acceptable time and stabilization of concomitant medication (4 weeks). Patients are then randomized to treatment groups including placebo and typically 2-3 doses of active study medication. A treatment phase of variable length (1-12 months) is initiated based on the characteristics of the drug and the mechanism of action. During the active treatment period, background medication must remain as stable as possible. The follow-up IVUS procedures are performed from 6 weeks to 12 months (see below for longer term IVUS studies) following the baseline IVUS.

II.2.i.F. Planned sample
Assuming that the standard deviation of the % change in plaque volume is around 6%, a sample size of approximately 65 patients per treatment group will provide 80% power to detect a 3% difference in plaque volume between the placebo group and one of the active treatment groups using a two-sided significance level of 0.05. Approximately 50 patients per treatment group are needed to detect a difference between the placebo group and two combined active groups in a 3-arm trial given the same assumptions. The final sample size must also account for loss-to-follow up of approximately 20 – 25%, given the cardiac catheterization procedures required.

II.2.i.G. Study population
Adults who require diagnostic coronary angiography for clinical indication, and stabilized to other background medications according to standard of care practices.

II.2.i.H. Specific inclusion criteria
a. Male or female aged ≥18 years
b. Female patients who are not of childbearing potential (at least two years postmenopausal, surgically sterile, or practicing adequate contraception)
c. Scheduled for clinically indicated coronary angiography (with or without PCI)
d. Presence of at least one luminal diameter stenosis of 20% or more in one coronary artery by visual (angiographic) estimation
e. Presence of a non-PCI target coronary artery in which IVUS examination can be performed (target vessel)
f. The target vessel must not have undergone previous PCI nor be a candidate for intervention at the time of the baseline catheterization

II.2.i.I. Specific exclusion criteria
a. Previous or planned coronary artery bypass surgery
b. Any medical or psychiatric condition which in the opinion of the investigator makes the patient an unsuitable candidate for the study

c. History of alcohol or drug abuse within the past year

II.2.i.J. Tools for assessing primary endpoints
Assessment of the main outcome measures of atherosclerosis regression is performed by 3-D reconstruction of IVUS images. All the IVUS images should be interpreted by experienced technicians supervised by a cardiologist blinded to treatment assignment. The baseline and end-of-treatment studies should be viewed together. The use of reproducible IVUS landmarks (i.e. aorto-ostial junction, branches) and a known pullback speed (0.5 mm/sec) facilitate comparison of the same 30-mm segment on both studies and permit volumetric (3-D) analysis. Frame-by-frame review of the images is also systematically used to confirm matching of segments. The images are digitized and quantitative analysis performed. The lumen and external elastic membrane (EEM) borders can be traced manually or using an edge detection algorithm if all tracings are visually verified.

Plaque, lumen and total vessel volumes are computed by multiplying the corresponding areas of each of the cross-section by the distance between neighbouring slices and by then adding all the products. Cross-sections are analyzed in the 30-mm segment of interest at both baseline and follow-up. Plaque, lumen and total vessel volumes are first computed for the entire length (30 mm) of the analyzed segment. Volumes are also calculated on 5-mm segments centered on the sites with a) smallest plaque burden at baseline and b) largest plaque burden at baseline. In addition to the absolute plaque volume, percent plaque volume is also calculated as plaque volume divided by total vessel volume times 100. Detailed analysis of plaque composition is then also performed with IVUS at baseline and follow-up using plaque characterization indexes.

For the quantitative analysis of coronary angiograms (QCA), all angiograms from a given patient should be viewed together and analyzed by experienced technicians supervised by a cardiovascular radiologist blinded to the patients’ treatment assignments.

II.2.i.K. Specific criteria for early withdrawal and discontinuation
Standard criteria for early withdrawal include withdrawal of consent, or adverse events that render the continued treatment of patient(s) medically unacceptable.

II.2.i.L. Data analysis method
The data analysis methods used may be dependent on the objective of the trial. One method is to base the analysis of efficacy variables on the intention-to-treat (ITT) population (including all randomized patients who have follow-up data). However, for exploratory purposes, the per-protocol population (patients with no major protocol violations) may also be investigated. Statistical tests are generally two-sided and p values ≤0.05 are often considered statistically significant. Percent change is analyzed using analyses of variance (ANOVA) models. Analyses of covariance (ANCOVA) models adjusting for baseline values are used for absolute change. Coefficients of correlation are computed to assess the relationship between changes in inflammatory markers and primary and secondary endpoints. The safety analysis includes adverse events presented by treatment group using descriptive statistics (frequencies and counts).

II.2.ii. Adjunctive therapy trial in patients with stable atherosclerosis – brachial artery reactivity

II.2.ii.A. Objectives
To evaluate short-term efficacy and tolerability during adjunctive therapy use

II.2.ii.B. Primary endpoints
a. Change in flow mediated dilation (FMD) defined as the percent change from brachial artery diameter (mm) at rest to maximal brachial artery diameter during reactive hyperemia.
II.2.ii.C. Secondary endpoints
Secondary endpoints are designed to distinguish between vasodilatory changes resulting from changes in endothelial function as opposed to smooth muscle function.
   a. Change in nitroglycerin-mediated dilation (percent change from brachial artery diameter before administration to maximal brachial artery diameter after administration of sublingual nitroglycerin, an endothelium-independent vasodilator)
   b. Change from baseline in normalized brachial artery diameter
   c. Change from baseline in blood pressure, heart rate and velocity versus time integral

II.2.ii.D. Exploratory endpoints
   a. Changes from baseline in markers of biochemical target activation and markers of atherosclerosis pathophysiology (i.e. inflammatory markers).

II.2.ii.E. Study design
A randomized, placebo-controlled, parallel-group design is generally used. Single-center or small multi-center studies are often preferable due to inter-operator variability. There are typically three phases, a phase for stabilization of concomitant medication and diet, a single-blind placebo-baseline phase, and a double-blind treatment phase. Inclusion criteria usually require patients to be at least 18 years of age with objective evidence of coronary artery disease, peripheral artery disease or carotid artery disease. At screening, patients are asked to refrain from significant changes in dietary, smoking, and exercise habits for the duration of the trial as these may affect their endothelial function. Following the baseline period (typically 4 weeks), patients have vascular reactivity assessed through ultrasound procedures performed on the brachial artery. Patients are then randomized to treatment groups including placebo and typically 2-3 doses of active study medication. A treatment phase of 8-12 weeks is initiated based on the characteristics of the drug and the mechanism of action. During the active treatment period, background medication must remain as stable as possible. The follow-up brachial assessments are performed at the end of the treatment period.

II.2.ii.F. Planned sample
A sample size of approximately 45 patients per treatment group will provide 80% power to detect a 2.5% difference in change in FMD between the placebo group and one of the active treatment groups assuming that the standard deviation of the change is 4% and using a two-sided significance level of 0.05.

II.2.ii.G. Study population
Adults with objective evidence of coronary artery disease, peripheral artery disease or carotid artery disease and stabilized to other background medications according to standard of care practices.

II.2.ii.H. Specific inclusion criteria
   a. Males or females aged ≥18 years
   b. Female patients who are not of childbearing potential (at least two years postmenopausal)
   c. Other specific inclusion criteria are based on defining objective evidence of atherosclerotic disease as follows:
   - Stable angina pectoris for which frequency, severity, duration, time of appearance, and precipitating events have not changed for 60 days prior to screening;
   - Myocardial ischemia as evidenced by any of the following
     ✓ Stress ECG showing ischemic ST-segment response
     ✓ Stress echocardiography showing myocardial wall motion abnormality, or
     ✓ Myocardial perfusion scan showing a myocardial perfusion defect
   - At least 50% occlusion of the lumen of one or more coronary arteries, as evidenced by coronary angiography, IVUS, or other methods
   - ECG evidence of Q-wave myocardial infarction
• Iliac, femoral or carotid artery atherosclerosis as evidenced by angiography, ultrasound duplex scan, IVUS or other methods
• Previous carotid endarterectomy, peripheral bypass surgery, or abdominal aneurysm.

II.2.ii.I. Specific exclusion criteria
a. History of myocardial infarction or unstable angina within 4 weeks prior to screening
b. History of PCI, coronary artery bypass, cerebrovascular accident, or diagnosis of heart failure within 3 months prior to screening
c. Uncontrolled hypertension
d. Uncontrolled diabetes
e. Concomitant administration of supplemental antioxidants, L-arginine supplements, dipyridamole, pentoxifylline, angiotensin-converting enzyme (ACE) inhibitors, or ACE receptor inhibitors.
f. Any medical or psychiatric condition which in the opinion of the investigator makes the patient an unsuitable candidate for the study
g. History of alcohol or drug abuse within the past year

II.2.ii.J. Tools for assessing primary endpoints
High resolution B-mode ultrasound imaging is used to determine the diameter of the right brachial artery before and immediately after reactive dilation induced by ischemia which is produced by inflating a blood pressure cuff to 200 mmHg or to 50 mmHg greater than the systolic blood pressure (whichever is higher) for 5 minutes. Arterial diameter at baseline and at 60 seconds after cuff deflation is recorded. The relative difference is the measure of reactivity. Image analysis should be performed by a central laboratory with experienced technicians blinded to study treatment.

II.2.ii.K. Specific criteria for early withdrawal and discontinuation
The investigator will discontinue a patient before completion of the study if consent is withdrawn or if it is medically unacceptable to continue treatment due to adverse events.

II.2.ii.L. Data analysis method
The analysis of efficacy variables may be based on the intention-to-treat (ITT) population (including all randomized patients who have follow-up data). All statistical tests are two-sided and p values ≤0.05 are considered statistically significant. Analysis of variance or analysis of covariance adjusting for the baseline values could be used to compare the primary and secondary endpoints between groups. In any case, data analysis methods should be planned a priori and chosen to answer the specific objectives of the trial.

II.3. Long-term studies
Since the accumulation of lipid rich plaque material develops slowly over several decades, it has generally been accepted that chronic therapy would be required in order to detect the effects of these agents on measures of plaque burden or subsequent risk of cardiovascular events. For example, while some recent IVUS trials have demonstrated efficacy of exploratory agents when administered over a period of weeks, approved lipid-lowering agents with a demonstrated ability to decrease cardiovascular morbidity and mortality have generally required treatment durations of 18 months or more in order to demonstrate plaque volume changes detectable with IVUS or carotid ultrasound. Therefore, trials may need to be designed to evaluate long-term administration in order to detect significant treatment effects. The primary objective of these long-term studies is to provide an opportunity for the drug effect to be realized and to obtain data on tolerability and safety during long-term use. The studies described above as short-term studies may also be considered long-term studies depending on the study drug mechanism of action and the anticipated time required for treatment effect. In addition to the use of QCA and IVUS in Phase II long-term studies to detect anti-atherosclerotic effects, carotid imaging is also an effective imaging tool.
II.3.i. Adjunctive therapy trial in patients with clinical indication for cardiac catheterization - IVUS

II.3.i.A. Objective
To evaluate long-term efficacy and tolerability during adjunctive therapy use

Study description is similar to short-term study description, except that long-term studies using IVUS generally provide for follow-up evaluations after 12-24 months.

II.3.ii. Adjunctive therapy trial in patients with stable atherosclerosis – carotid intima-media thickness (IMT) evaluations

II.3.ii.A. Objective
To evaluate long-term efficacy and tolerability during adjunctive therapy use

II.3.ii.B. Primary endpoints
a. The change over time in mean maximum IMT across 12 pre-selected carotid arterial segments.

II.3.ii.C. Secondary endpoints
a. The difference in the slope of left and right common carotid artery far wall mean IMT progression (up to 200 far-wall measurements over a distance of approximately 2 cm beginning 1 cm below the bifurcation and using a standard angle of imaging)
b. The change in mean far wall IMT
c. The change in maximum far wall IMT
d. The rate of carotid artery progression measured as linear slope over annual ultrasound examinations
e. The average of the maximum carotid IMT of the far wall of up to 4 arterial segments (right and left distal common and right and left carotid bulb)
f. The rate of progression in the far wall of the left and right carotid bulb IMT
g. The rate of progression of the mean of maximum IMTs of the left and right common carotid artery
h. Change from baseline in mean IMT of the left and right common carotid arteries
i. Change from baseline in the maximum IMT of the left and right common carotid arteries
j. Incidence of adverse events.

II.3.ii.D. Exploratory endpoints
Changes from baseline in markers of biochemical target activation and markers of atherosclerosis pathophysiology (i.e. inflammatory markers)

II.3.ii.E. Study design
A multicenter, randomized, placebo-controlled, parallel-group design is generally used. Inclusion criteria usually require patients to be at least 18 years of age with evidence of carotid and coronary atherosclerosis. Patients are usually required to have stable background treatment following standard care recommendations. Patients are screened and the initial carotid ultrasound examination is performed bilaterally on the common and internal arteries. Patients are then randomized to treatment groups including placebo and typically 2-3 doses of active study medication. It is generally thought that treatment effects on the common carotid artery mean IMT could be seen after one year of treatment, but an 18-24 month treatment period would enhance the likelihood of this. Due to the non-invasive nature of carotid IMT assessment, serial examinations of the carotid wall are often undertaken.

II.3.ii.F. Planned sample
The sample size calculation is based on an average expected effect of 0.02 mm/yr change from baseline in carotid IMT measurements in at least one treatment group (for a total change of 0.04 mm in a 24-month
study). No change is expected in the placebo group. According to the literature, 0.20 mm would be a reasonable estimate of the standard deviation of the change in IMT from baseline to 24 months. Under these assumptions, a sample size of approximately 400 patients per treatment group would provide 80% power to detect a difference of 0.04 mm in change in IMT between the placebo group and one of the active treatment groups using a two-sided significance level of 0.05.

II.3.ii.G. Study population
Patients with evidence of atherosclerotic disease

II.3.ii.H. Specific inclusion criteria
a. Patients may be selected based on a pre-specified minimal baseline IMT measurement (> 0.8 mm),
b. Patients may be selected based on other evidence of atherosclerotic cardiovascular disease
c. Patients may be selected based on both of these criteria.

II.3.ii.I. Specific exclusion criteria
a. History of carotid revascularization
b. Patients in whom a screening IMT is suboptimal
c. Any medical or psychiatric condition which in the opinion of the investigator makes the patient an unsuitable candidate for the study
d. History of alcohol or drug abuse within the past year

II.3.ii.J. Tools for assessing primary endpoints
Images of the right and left common carotid and internal carotid arteries are captured, including images of the near and far wall, using high-resolution B-mode ultrasound. Ultrasound methodology should be specifically designed to include procedures to quality control the critical components of measurement variation including instrumentation, and ultrasound operations. Standardization of ultrasound machines at all sites is optimal, but not necessary. Image analysis is performed centrally at a center with experienced technicians.

II.3.ii.K. Data analysis method
The analysis of efficacy variables may be based on the intention-to-treat (ITT) population (including all randomized patients who have follow-up data). The continuous efficacy endpoints are analyzed using an analysis of covariance (ANCOVA) model with treatment and center as effect and the parameter’s baseline as covariate. If serial examinations of the carotid wall are undertaken, repeated measures ANOVA could also be conducted to study the way IMT changes over time. Chi-square tests or Cochran-Mantel-Haenszel tests using center as a stratification factor are used for categorical endpoints. Other data analysis methods more suitable for exploratory evaluations may be employed, depending on the objectives of the trial.

III. PHASE III STUDIES FOR REGISTRATION OF NEW ANTI-ATHEROSCLEROTIC DRUGS

III.1. Outline of a typical development plan
Phase III studies are conducted to establish the overall risk-benefit relationship of the drug and to provide adequate information for drug labeling. Therefore, for an atherosclerosis indication, the endpoints, patient population, and patient numbers should be consistent with these goals. It is important to recognize that no anti-atherosclerotic drugs working by a non-lipid mechanism have yet been approved. Additionally, there are no official guidelines written to direct the development of such agents. Three concerns have previously been identified for drugs working through novel, non-LDL cholesterol mechanisms. The first concern is that use of true placebo-controlled trials is not ethical in high-risk patients. Secondly, the administration of background “usual care” consisting of statins and other lipid-lowering therapy may add to the complexity of trial design and interpretation. Finally, with the need to demonstrate the effectiveness
of adjunctive therapy, sample size may become prohibitively large. The use of the imaging technologies outlined in Phase II trial descriptions may some day provide sufficient data to support the use of these endpoints as a surrogate for the reduction of cardiovascular risk. However, to date, the ability to rely solely on imaging endpoints in Phase III is yet theoretical. Therefore the development plan outlined in this document consists of a multicenter (usually multi-national), double-blind, randomized, placebo-controlled trial of two parallel groups designed to assess the combined incidence of cardiovascular morbidity and mortality.

### III.2. Long-term adjunct therapy trial

#### III.2.i. Adjunctive therapy trial in patients at high risk for a major cardiovascular event

#### III.2.i.A. Objectives

To assess the effect of the investigational drug versus placebo on the combined incidence of cardiovascular morbidity and mortality

The definition of cardiovascular morbidity may include only “hard” endpoints such as non-fatal myocardial infarction, and stroke, or be expanded to include other “soft” endpoints, such as the need for coronary revascularization procedures, worsening angina requiring hospitalization, objective evidence of ischemia, and incidence of peripheral arterial disease.

#### III.2.i.B. Primary endpoint

- Combined incidence of cardiovascular morbidity and mortality

#### III.2.i.C. Secondary endpoints

- Incidence of all cause mortality
- Incidence of cardiovascular mortality
- Incidence of cardiovascular morbidity
- Combined incidence of a subset of the events within the cardiovascular morbidity and mortality definitions.
- Incidence of adverse events.

#### III.2.i.D. Exploratory endpoints

Changes from baseline in markers of biochemical target activation and markers of atherosclerosis pathophysiology (ie. inflammatory markers).

#### III.2.i.E. Study design

A multicenter, randomized, adjunctive therapy, double-blind, placebo-controlled parallel-group design could be utilized. Patients should receive standard care, including treatment of underlying diabetes, hypercholesterolemia and hypertension, prior to entry into this trial. There are several reasonable approaches to overall study design. One approach is to design the trial so that it will be complete with a fixed number of patients completing a minimal, fixed treatment period. Another approach is to design the trial so that it will complete when a predetermined number of patients have experienced a primary event based on expected control group rates and anticipated treatment effects.

#### III.2.i.F. Planned sample

The sample size will be based on the expected event rate in the selected patient population as well as the anticipated magnitude of treatment effect as demonstrated in Phase II trials. Other factors to consider in determining sample size include standard of care in country/region where trial is being conducted, planned trial duration, and time to effect based on an individual agent’s mechanism of action. A sample size of 2,000
– 6,000 moderate to high-risk patients per treatment group with minimally 18-month follow up would be a reasonable expectation.

III.2.i.G. Study population
Patients at moderate to high risk of cardiovascular morbidity and mortality based on the presence of one or more risk factors. Patients should receive treatment for modifiable risk factors prior to entry into trial.

III.2.i.H. Specific inclusion criteria
a. Specific inclusion criteria are selected based on identification of patients with moderate to high risk of cardiovascular morbidity and/or mortality.
b. Inclusion criteria may include one or more of the following major risk factors:
   • Cigarette smoking
   • Hypertension (BP \( \geq 140/90 \) mm Hg)
   • Low HDL cholesterol \( < 40 \) mg/dL
   • Family history of premature CHD (CHD in male first degree relative \( < 55 \) years; CHD in female first degree relative \( < 65 \) years)
   • Age \( \geq 55 \) years
   • Diabetes
c. Inclusion criteria may include one or more of the following life-habit risk factors
   • Obesity
   • Physical inactivity
   • Atherogenic diet
d. Inclusion criteria may include one or more of the following emerging risk factors
   • C reactive protein
   • Lipoprotein (a)
   • Homocysteine
   • Prothrombotic and proinflammatory factors
   • Impaired fasting glucose
   • Evidence of subclinical atherosclerotic disease

III.2.i.I. Specific exclusion criteria
a. Any medical or psychiatric condition which in the opinion of the investigator makes the patient an unsuitable candidate for the study
b. History of alcohol or drug abuse within the past year

III.2.i.J. Tools for assessing primary endpoints
An independent Clinical Endpoint Committee (CEC) is established to review and classify all suspected major cardiovascular events according to pre-defined guidelines and definitions. The CEC remains blinded to treatment group assignment for the duration of the trial. The CEC is typically composed of 5 cardiologists experienced in patient care. Reviewers are assigned at random to review each case and classify the event. For the classification to be considered final, consensus between two primary reviewers is required. If consensus between two primary reviewers is not reached, the case is typically decided by simple majority of the entire committee.

III.2.i.K. Specific criteria for early withdrawal and discontinuation
The investigator will discontinue a patient before completion of the study if consent is withdrawn or if it is medically unacceptable to continue treatment due to adverse events. An independent Data Safety Monitoring Board (DSMB) is typically established and responsible for assessing patient safety during the course of the trial. A trial may be discontinued if such a recommendation is made by the DSMB based on a periodic review of the generated data.
III.2.i.L. Data analysis method
The analysis of efficacy variables is based on the intention-to-treat (ITT) population. Statistical tests are 2-
sided at the 0.05 level of significance. Survival analysis including Kaplan-Meier curves and log-rank tests to
compare survival curves across groups are used. Multivariate analysis using Cox proportional hazards
models may also be performed.

IV. OTHER STUDIES (SPECIAL INDICATIONS)
Other studies may be considered for inclusion in the development plan for anti-atherosclerotic agents that
work by non-lipid mechanisms. Patient populations that may warrant focused evaluation in separate
clinical trials include patients with traditional or emerging risk factors, or genetically defined conditions
which confer high risk of CAD. Separate trials in patients with diabetes, metabolic syndrome, genetic
predisposition to CAD, peripheral arterial disease, etc. may provide focused and useful information for
further development. There may be also other imaging technologies that may have an important role in
drug development aside from those mentioned previously. Magnetic resonance imaging, computed
tomography, positron emission tomography and other technologies may all have an important role in
determining the effectiveness of potential anti-atherosclerotic agents. These additional studies may have
utility in early stage development to help establish proof of concept and better define the mechanism of
action. Alternately further evaluations of specific populations and technologies may have greatest utility
in Phase IV trials to further define established efficacy.

V. EXAMPLES OF LANDMARK WELL DESIGNED TRIALS

Intravascular Ultrasound
1. Tardif JC, Gregoire J, Schwartz L, Title L, Laramee L, Reeves F, Lesperance J, Bourassa MG,
   L’Allier PL, Glass M, Lambert J, Guertin MC; Canadian Antioxidant Restenosis Trial (CART-1)
   Investigators. Effects of AGI-1067 and probucol after percutaneous coronary interventions.
2. Nissen SE, Tsunoda T, Tuzcu EM, Schoenhagen P, Cooper CJ, Yasin M, Eaton GM, Lauer MA,
   Sheldon WS, Grines CL, Halpern S, Crowe T, Blankenship JC, Kershner R. Effect of recombinant
   ApoA-I Milano on coronary atherosclerosis in patients with acute coronary syndromes: a randomized
   controlled trial. JAMA 2003;290:2292-2300.
   JF, Nattel S, Gallo R, Crepeau J. Effects of probucol on vascular remodeling after coronary angioplasty.
   Heinen T. Design features of the Avasimibe and Progression of coronary Lesions assessed by

Brachial artery reactivity
   Benefits of lipid lowering on vascular reactivity in patients with coronary artery disease and average
2. Dupuis J, Tardif JC, Cernacek P, Theroux P. Cholesterol reduction rapidly improves endothelial
   function after acute coronary syndromes. The RECIFE (reduction of cholesterol in ischemia and

Carotid intima-media thickness (IMT)
   treatment effect on carotid and femoral artery walls and its correlations with coronary arteriographic


**Major cardiovascular morbidity and mortality studies**


**VI. SUGGESTED READINGS**


Chapter 14. Drugs for Heart Failure

Anique Ducharme, M.D., M.Sc. ¹

Simon de Denus, B. Pharm, M.Sc. ²

¹ Montreal Heart Institute
5000 Belanger Street East,
Montreal, Quebec  HIT 1C8
CANADA
Tel 514-376-3330 ext.3927
Fax 514-593-2575
Email ducharme@icm.umontreal.ca

² Faculty of Pharmacy
University of Montreal
Montreal, Québec
CANADA
and
Department of Pharmacy
Montreal Heart Institute
Montreal, Québec
CANADA

Corresponding Author:
Anique Ducharme, M.D., M.Sc.
I. INTRODUCTORY REMARKS

Heart failure (HF) has reached epidemic proportions in the U.S. with nearly 5 millions individuals afflicted with this condition. Despite major advances in the treatment of this condition, both pharmacological and non-pharmacological, HF is responsible for more than 250 000 deaths per year in the US alone. Furthermore, this condition is associated with significant morbidity with almost 1 million hospital admissions each year. Because of this, the management of HF is associated with an annual cost of more than 24 billion dollars. These statistics clearly illustrates that novel therapies, both pharmacological and non-pharmacological, are still required to improve the prognosis of these patients. Unfortunately, the development of new HF drugs is complicated by a lack of established surrogate markers as reliable as blood pressure in hypertension or cholesterol concentrations in dyslipidemia.

II. PHASE II STUDY TO EVALUATE THE EFFECTIVENESS AND SAFETY OF NEW HEART FAILURE DRUGS

II.1. Outline of a typical development plan

This phase generally begins with open, single-blind or double-blind studies comparing various dosages of a new agent in addition to or compared to an established treatment. Frequently, these dosages are derived from the hypertension literature, because these dosages have been proven to be safe and effective. These studies are generally conducted in a limited number of centers.

Results of these pilot studies are vital for appropriate design of key studies, since these small initial trials then lead to large dose-finding multicenter, controlled trials. These studies are generally conducted using a parallel design and compare the investigational drug with placebo or an active-control, depending if it is expected to be used in addition to current treatment or as an alternative to an established treatment. For example, angiotensin receptor blockers (ARBs) were initially compared to angiotensin-converting enzyme (ACE) inhibitors because they were expected to provide a superior clinical benefit through a more complete inhibition of the effects of the renin-angiotensin-aldosterone system, while being better tolerated because they do not cause an accumulation of bradykinin. Oppositely, when experimental data demonstrated that bradykinin accumulation could positively influence the evolution of HF, the addition of an ARB in patients already receiving an ACE inhibitor was investigated with positive results. These dose-finding studies will generally involve patients with HF of moderate to severe symptoms of New York Heart Association (NYHA) II-IV. Asymptomatic patients are usually excluded (NYHA functional class I). End points commonly consist of vital signs, exercise tolerance, hemodynamic measurements, echocardiographic measurements, neurohormone concentrations and patient’s assessment of well-being.

The potential for drug interactions with commonly used cardiovascular drugs with a high potential for interaction (warfarin, digoxin) and other drugs known to affect the metabolism of commonly used drugs (cimetidine, rifampin) may also be investigated, but these investigations are generally conducted in healthy or hypertensive patients and, unfortunately, only rarely in HF patients. Because the pharmacokinetics of some agents vary between healthy individuals and patients with HF, it is possible that the importance of a drug interaction could vary between these two patient populations and it would be therefore preferable to evaluate drug interactions of interest in the HF patient population also.

II.2. Short-term studies

II.2.A. Objectives
Evaluate both the short-term efficacy, generally through hemodynamic monitoring, neurohormonal measurements and functional class, and the safety of the agent.
II.2.B. Primary endpoints
Because there is no established surrogate marker to evaluate the efficacy of drugs in HF, endpoints vary greatly in short term studies, but generally consist of either:
   a. Hemodynamic measurements (filling pressures, cardiac index, systemic vascular resistance, blood pressure, heart rate)
   b. Improvement in symptoms or functional class
   c. Improvement in exercise capacity (peak oxygen consumption, sub maximal exercise tolerance or 6 minutes walk test).
   d. Measurements of neurohormones: catecholamines, brain natriuretic peptide (BNP), others.
   e. Echocardiographic or Magnetic Resonance Imaging (MRI) parameters: left ventricular (LV) dimensions, volumes, LV ejection fraction (LVEF), and degree of mitral regurgitation.
   f. Patients’ perception of dyspnoea and well-being.

II.2.C. Secondary endpoints
In these initial phase II studies, secondary endpoints will consist of the incidence of adverse drug effects (safety and tolerability) and mortality, cardiovascular events (CV mortality, myocardial infarction, stroke and cause-specific hospitalizations). The impact of the agent under investigation on quality of life may also be evaluated.

II.2.D. Exploratory endpoints
Relationship between dose, efficacy, and adverse effects. Drug withdrawal and rebound effects.

II.2.E. Study design
A least one randomized, double-blinded, parallel-group design, using multiple centers is generally conducted. Patients can be randomized to different doses of the investigated drug versus placebo or an active control, usually with a dose-ranging design. Patients should be receiving an optimal medical therapy at proven or maximally tolerated doses at the beginning of the study; they should be maintained on this regimen throughout the duration of the study. The drug titration period will depend on the agent being investigated and the need to monitor for vital signs, electrolytes, renal and liver function tests. Maintenance phase should be of at least 3 months duration, but no longer than 6 months. At the end of the study, the patients should be withdrawn from the investigational drug based on a pre-established schedule. If the drug is planned to be on the market in a near future, an open label extension phase can be offered to the patient.

Concomitant therapy
Unless contraindicated, all patients should be on optimal medical therapy (both agents and doses) consisting of agents from the following class of agents:
   a. ACE inhibitors or ARBs
   b. Beta-blockers
   c. Spironolactone (in NYHA class III-IV patients)
   d. Diuretics, usually
   e. Digoxin

II.2.F. Planned sample
In these initial trials, the sample size will depend on the end points measured. Typically the smaller initial trials will include approximately 50 to 100 patients, whereas the larger dose-finding trial can include up to a few hundred patients.

II.2.G. Study population
Patients with symptomatic, chronic HF
II.2.H. Specific inclusion criteria
   a. Adults (at least 18 years of age)
   b. Left ventricular ejection fraction ≤40% (or 35%), measured in the last 6 months.
   c. NYHA class II-IV symptoms.
   d. Stable symptoms: duration of stability varies according to study (4 days-3 months, except studies of decompensated HF)
   e. Stable medical regimen ≥1 month (except diuretics)

II.2.I. Specific exclusion criteria
   a. Inability to provide informed consent
   b. Unstable angina, myocardial infarction or coronary revascularisation within the last 6 weeks-3 months.
   c. Planned cardiac surgery (within 3-6 months).
   d. Significant valvular disease.
   e. Systolic blood pressure < 90 mmHg (except in studies of cardiogenic shock).
   f. Uncontrolled hypertension (blood pressure systolic >170 mm Hg or diastolic >105 mm Hg).
   g. Heart rate < 60 bpm for agent causing significant bradycardia.
   h. Advanced heart block without pacemaker.
   i. Significant renal insufficiency (definition varies between trials and the nature of the drug studied).
   j. Significant non-cardiac co-morbidities or laboratory abnormalities.
   k. Pregnancy and women of childbearing potential unless a safe contraception method is used.
   l. A potentially reversible cause of HF (e.g. thyrotoxicosis or uncontrolled supraventricular arrhythmia).
   m. Known drug or alcohol misuse, poor compliance with treatment or any other serious systemic disease that might complicate management and reduce life expectancy.
   n. Administration of any investigational drug within the preceding 30 days.

II.2.J. Tools for assessing endpoints
Primary endpoints
   a. Echocardiography or MRI
   b. Exercise stress test with or without O₂ consumption measurements
   c. Right heart catheter
   d. Blood samples
   e. Visual analog scale (for symptoms’ self-evaluation)

Secondary endpoints
   a. Emergency room visits log
   b. Discharge summary: cause & number of hospital admission
   c. Death certificate
   d. Quality of life questionnaire

The measure of the endpoints will be performed according:
   a. Review of the Case Report forms for all clinical events, including queries of the patient’s clinical chart if needed
   b. Standardized method of evaluation: Core laboratory for consistency of data: echocardiography, MRI, blood samples (neuro-hormones, BNP, etc.), and exercise stress test.

II.2.K. Specific criteria for early withdrawal and discontinuation
   a. Patient’s request/consent withdrawal
   b. Pregnancy
c. Serious adverse effects
d. When, in the opinion of the physician, continuation of the therapy is not in the best interest of the patient.

II.2.1. Data analysis method
The analysis of efficacy variables is based on the intention-to-treat principle, which includes all randomized patients. The specific statistical analyses will vary according to the types of variables measured and endpoints evaluated. All statistical tests are two-sided and statistical significance is generally established with a p value ≤ 0.05. Multiple statistical methods can be used for the primary and secondary end points.

II.3. Long-term studies
At the end of the randomized phase, it is usual for patients completing short-term studies to be offered an extension phase with open-label follow-up. The objective of this prolongation is to provide data on tolerability and safety during long-term use. The study drug may be continued for as long as it is felt to be clinically beneficial, until the agent is available on the market.

III. PHASE III STUDY TO EVALUATE THE EFFECTIVENESS AND SAFETY OF NEW DRUGS FOR HEART FAILURE

III.1. Outline of a typical development plan
Whereas phase II studies generally focus on surrogate markers of efficacy, phase III trials are specifically designed to evaluate the impact of the drug under investigation on clinical outcomes, mainly death and hospitalization. This phase generally includes at least one large multicenter double-blind, randomized placebo or active control trial of patients with symptomatic HF (NYHA functional class II-IV). Patients with NYHA functional class I are generally excluded because of their lower risk of cardiovascular events. For patients in NYHA II, a criteria for potential instability is often required (ex. hospital admission for HF within 6-12 months or use of diuretics). The follow-up of these studies should be of at least 6 months, preferably 12 months, but can be extended to several years.

III.2. Short-term studies
The design for phase III trials in HF is similar to that described above for phase II trials. Sometimes, multiple doses or regimens are explored.

III.3. Long-term studies

III.3.A. Objectives
To evaluate the long-term efficacy and safety of a given treatment compared with optimal medical management.

III.3.B. Primary endpoints
Mortality (total, cardiovascular), or a combination of mortality and morbidity (which generally consists of cardiovascular hospitalizations or episodes of worsening HF) is usually used.

III.3.C. Secondary endpoints
Secondary endpoints for phase III trials will consist of any of the previously mentioned endpoints not included in the primary endpoints or any components of a combined end point. In addition, drug safety and tolerance is reported as a secondary endpoint. Furthermore, the following can also constitute secondary endpoints: ischemic events (myocardial infarction, angina, need for revascularisation), changes in neurohormones or inflammatory markers levels, exercise tolerance, quality of life or changes in
echocardiographic (or MRI) parameters (left ventricular dimensions, LV ejection fraction [LVEF], ventricular volumes, degree of mitral regurgitation, etc…).

III.3.D. Exploratory endpoints
Exploratory analyses are generally limited to any sub-group analysis not prospectively defined.

III.3.E. Study design
Phase III study are usually multicenter, randomized, double-blinded placebo-controlled or active-controlled studies.

Concomitant therapy
Optimal medical treatment as described for the phase II in section II.2.E.

III.3.F. Planned sample
In phase III trials, the sample size is calculated based on the number of events expected in the population studied during the course of the trial and the expected reduction on these events rate with the new treatment. Usually, several thousand patients are planned to be enrolled.

III.3.G. Study population
As described for phase II studies in the section II.2.G.

III.3.H. Specific inclusion criteria
The inclusion criteria are generally the same as those used for the phase II.

III.3.I. Specific exclusion criteria
The exclusion criteria are generally the same as those used for the phase II. In addition, patients awaiting a cardiac transplantation are often excluded.

III.3.J. Tools for assessing endpoints
Primary endpoints
   a. Clinical end-points committee (CEC): an independent CEC is usually formed to review every event occurring during the course of the study, with blind adjudication.

Secondary end points
   a. Discharge summary (cause & duration of hospitalization)
   b. Worsening HF: chest X-Ray, BNP level, IV diuretic used or number of unplanned HF clinic visits.

The measurement of endpoints is performed as described for the phase II studies in the section II.2.J.

III.3.K. Specific criteria for early withdrawal and discontinuation
These criteria are generally the same as those discussed for phase II trials in section II.2.K.

III.3.L. Data analysis method
The data analysis is performed in an intention to treat fashion although an “on treatment” analysis is sometimes performed retrospectively in some cases (e.g. high rate of withdrawal).

IV. PHASE III STUDY TO EVALUATE THE EFFECTIVENESS AND SAFETY OF NEW DRUGS FOR HEART FAILURE IN SPECIAL POPULATIONS

The majority of HF trials have previously focused on patients with depressed left ventricular systolic function (ex. LVEF < 40%), probably because they have a poorer prognosis than patients with symptomatic HF and
preserved systolic function. Nevertheless, this preserved HF group represents almost a third of all HF patients, but has not been studied extensively. With the aging of the population and the prevalence of diabetes, which has reached epidemic proportions, this group will increase dramatically. Thus, efforts have been put forward lately to study more carefully HF patients with preserved systolic function. Unfortunately, the diagnosis is often difficult, since dyspnoea is often multifactorial (including participation of cardiac causes, hypertension and pulmonary disease). Furthermore, there is a lack of consensus on the criteria to define and investigate this entity. Fortunately, a potential surrogate marker could help differentiate between shortness of breath of pulmonary or cardiac origin, the BNP, which is a hormone secreted mainly by the ventricle in response to stretch.

**IV.1. Outline of a typical development plan**

Similar to phase III study as outlined in section III.1, except for the inclusion criteria.

**IV.2. Long term studies**

**IV.2.A. Objectives**
As described for phase III studies in section III.3.A.

**IV.2.B. Primary endpoints**
As described for phase III studies in section III.3.B.

**IV.2.C. Secondary endpoints**
As described for phase III studies in section III.3.C.

**IV.2.D. Study design**
As described for phase III studies in section III.3.E.

Concomitant therapy
Optimal medical treatment is unknown in this patients’ group, and usually relies on the treatment of concomitant illnesses (ex. ischemia, dyslipidemia, diabetes, hypertension, etc.).

**IV.2.E. Planned sample**
As for phase III trials of HF patients with systolic dysfunction, the sample size is calculated based on the number of events expected in the population studied during the course of the study and the expected reduction on these events rate with the new treatment. Unfortunately, since HF with preserved systolic function has not been fully studied, assumptions on prognosis and number of events are currently difficult to assess.

**IV.2.F. Study population**
- Preserved LV systolic function (EF > 40%)
- Others, as outline for phase II studies in section II.2.G.

**IV.2.G. Specific inclusion criteria**
The inclusion criteria are similar to those described for phase II studies in section II.2.H.

**IV.2.H. Specific exclusion criteria**
The exclusion criteria are similar to those described for phase II studies in section II.2.I.

**IV.2.I. Tools for assessing endpoints**
The tools to assess and measure the endpoints are usually the same as those used during the phase III in section III.3.J.
IV.2.J. Specific criteria for early withdrawal and discontinuation
Usually the same as those discussed for phase II trials in section II.2.K.

IV.2.K. Data analysis method
Usually the same as those used for the phase III in section III.3.L., but special emphasis should be put on the number of events needed (see above).

V. PHASE IV STUDIES
Phase IV studies are usually large-scale, open-label postmarketing surveillance trials. These studies take place once the FDA or EMEA has approved the agent. The goal is to evaluate if the safety and efficacy of the agent evaluated in controlled studies is maintained in a large population. These usually include patients from subgroups potentially at higher risk for adverse events. Physicians systematically document their observations concerning patients with HF on case report forms. This trial design is rarely used in the HF population.

V.1. Outline of a typical development plan
Similar to phase II and III studies outlined above.

V.2. Short-term study

V.2.A. Objectives
The goal is to find if the safety and efficacy of the pharmacological agent as evaluated in controlled studies is maintained in a large population. The duration is usually 6 months.

V.2.B. Primary endpoints
Similar to phase II and III studies outlined in section II.2.B. and III.3.B. In addition, global tolerability is rated among specific subgroups of patients: genders, age and co-morbidities. Quality of life questionnaire can also be included.

V.2.C. Secondary endpoints
Similar to phase II and III studies outlined in sections II.2.C. and III.3.C.

V.2.D. Study design
Randomized, active-controlled, open-label study

Concomitant therapy
As described in the phase II outline in section II.2.E.

V.2.E. Planned sample
While for antihypertensive drugs several thousand patients are usually necessary, the study of drugs indicated for HF usually include smaller cohorts. The number of patients enrolled in these trials have been as small as 50 and, generally, is limited to no more than a few hundred patients.

V.2.F. Study population
As described in the phase II studies outlined in section II.2.G.

V.2.G. Specific inclusion criteria
As described in the phase II studies outlined in section II.2.H.
V.2.H. Specific exclusion criteria
As described in the phase II studies outlined in section II.2.I.

V.2.I. Tools for assessing endpoints
a. Tolerability questionnaire
b. Blood samples
c. Exercise tolerance test

Measure of endpoints
As described in the phase II outlined in section II.2.J.

V.2.J. Specific criteria for early withdrawal and discontinuation
As described in the phase II outlined in section II.2.K.

V.2.K. Data analysis method
As described in the phase II outlined in section II.2.L.

V.3. Long-term studies

V.3.A. Objectives
The goal is to compare the effects of drug A and drug B on clinical outcome. The duration may be up to 5 years. The only example of a phase IV study in patients with HF is the recently published COMET (Carvedilol Or Metoprolol European Trial). Since beta-blockers have been shown to reduce mortality in HF patients with systolic dysfunction, they aimed to compare the effects of carvedilol and metoprolol on clinical outcome, in patients who were on background treatment with diuretics and ACE inhibitors. The methods of the COMET trial will be used as an example of phase IV study.

V.3.B. Primary endpoints
The primary end points of the COMET study were all-cause mortality and the composite end point of all-cause mortality or all-cause admission.

V.3.C. Secondary endpoints
Similar to the ones described for phase III trial.

V.3.D. Study design
Randomized, active-controlled, open-label or double-blind study (COMET).

Concomitant therapy
a. On stable HF treatment with ACE inhibitors for $\geq 4$ weeks unless contraindicated.
b. On treatment with diuretics ($\geq 40$ mg of furosemide or equivalent) for at least 2 weeks.
c. Digitalis, ARBs, or other vasodilators could be used at the discretion of the investigators.

V.3.E. Planned sample
The COMET study was planned as an event-driven parallel-group survival study to compare carvedilol and metoprolol with respect to all-cause mortality. A total of 1020 fatal events were needed to detect a risk reduction of 20% with at least 80% power with an overall type I error of 0.05.

V.3.F. Study population
Eligible patients were men or women with symptomatic chronic HF (New York Heart Association [NYHA] class II–IV),
V.3.G. Specific inclusion criteria
   a. At least one cardiovascular admission during the previous 2 years.
   b. LV ejection fraction ≤ 0.35, measured within the previous 3 months or left-ventricular end diastolic diameter ≥ 6.0 cm and a fractional shortening < 20% measured by echocardiography.

V.3.H. Specific exclusion criteria
   a. A recent change of treatment, current treatment with diltiazem or verapamil, amiodarone (>200 mg per day) or class-I antiarrhythmic drugs.
   b. Patients with unstable angina, myocardial infarction, or coronary revascularisation or stroke within the previous 2 months.
   c. Uncontrolled hypertension (blood pressure systolic > 170 mm Hg or diastolic > 105 mm Hg).
   d. Hemodynamically significant valvular disease.
   e. Symptomatic and sustained ventricular arrhythmias within the past 2 months not adequately treated with antiarrhythmic drugs or defibrillator.
   f. Pregnancy, women with childbearing potential on inadequate contraception.
   g. Known drug or alcohol misuse, poor compliance with treatment.
   h. Any other serious systemic disease that might complicate management and reduce life expectancy.
   i. Patients in whom there was a contraindication to use of β-blockers, such as resting heart rate< 60 bpm, sick sinus syndrome, bifascicular block, second or third degree atrioventricular block unless treated with a pacemaker, sitting systolic blood pressure < 85 mm Hg, history of asthma or chronic obstructive pulmonary disease, peripheral arterial disease with symptoms at rest, or unstable insulin-dependent diabetes mellitus.

V.3.I. Tools for assessing primary endpoints
As described in the phase III outlined in section III.3.J.

Measure of end points
As described in the phase III outlined in section III.3.J.

V.3.J. Specific criteria for early withdrawal and discontinuation
As described in the phase II outlined in section II.2.K.

V.3.K. Data analysis method
Analysis was done by intention to treat.

V.4. Comments
Although the phase IV study design is common in clinical research focusing on hypertension, it is infrequently conducted in the HF population. Only one large phase IV study including patients with HF has been published, which is described in detail above.

VI. EXAMPLES OF LANDMARK WELL DESIGNED TRIALS

Phase II studies
Phase III studies

Phase III –Special population (HF with preserved LVEF)

Phase IV

VII. SUGGESTED READINGS
Chapter 15. Arrhythmias

Teresa Kus, M.D., Ph.D.

Associate Professor
Department of Pharmacology
Université de Montréal
Centre de Recherche
Hôpital du Sacré Cœur
5400, boulevard Gouin Ouest
Montréal, Québec H4J-1C5
CANADA
I. INTRODUCTORY REMARKS

Historical Perspective
The development of new antiarrhythmic drugs has become a major challenge since the publication of the Cardiac Arrhythmia Suppression Trial (CAST) results in 1989 (1). This was the first study designed to demonstrate that antiarrhythmic drugs with a potent Class I effect (flecainide and encainide) decrease mortality as compared to placebo in patients with remote myocardial infarction (MI) and frequent ventricular ectopic beats and who are therefore at increased risk for sudden death (SD) due to malignant ventricular arrhythmia. Improved survival was expected as these drugs markedly suppressed premature ventricular contractions in the same study patients. Instead, an increase in mortality was observed on antiarrhythmic drug compared to placebo (2,3).

This result affected antiarrhythmic drug development in two ways. First, the search for more potent drugs with Class I effect, i.e., slowing of conduction velocity by blockade of the sodium channel, was abandoned. In the setting of prior MI or ongoing ischemia, Class I effect was now generally accepted as potentially proarrhythmic. Instead, attention was directed to drugs which prolong action potential duration and refractoriness. In the presence of such a Class III effect, an arrhythmia dependent on a rapidly circulating electrical impulse would be interrupted when the wavefront encountered refractory tissue. The second major result of CAST was the perceived necessity of demonstrating the safety of such drugs in large placebo-controlled trials in patients with underlying cardiac pathology and left ventricular (LV) systolic dysfunction. This also proved to be important for the Class III drugs (3). In susceptible populations or under particular clinical circumstances, these drugs can induce excessive and heterogeneous prolongation of action potential duration with the consequent appearance of triggered automaticity. Such abnormalities can then set the conditions for torsade des pointes, a potentially lethal ventricular arrhythmia (4).

Present Indications for Antiarrhythmic Drug Therapy
Currently, the major indications for antiarrhythmic drug use are atrial fibrillation (AF) and atrial flutter (AFl). AF is a rapid (>400 bpm), disorganized, multi-circuit atrial rhythm whose prevalence increases with age (up to 8% in those over 80 years). It is responsible for approximately one-third of arrhythmia admissions and increases mortality two-fold and the risk of stroke five-fold (5). Acutely, it can manifest with symptoms of palpitations, hypotension, angina or heart failure. Further, if the ventricular response rate remains chronically elevated, it can result in a tachycardia-induced cardiomyopathy (CM). AFl, a single organized circuit usually in the right atrium, is less common but also increases the risk of cardiogenic emboli and provokes similar symptoms. Antiarrhythmic drugs are used acutely as an alternative to electrical cardioversion for both persistent AF and AFl so as to achieve sinus rhythm (SR) and chronically to prevent recurrence. They are also used to diminish the frequency of paroxysmal or spontaneously terminating episodes of AF and AFl. The currently available antiarrhythmics, however, are only moderately effective (30-60%), only a few are safe to administer in the presence of heart failure (dofetilide, amiodarone), and, in the case of the most effective (amiodarone), risk significant organ toxicity (6).

The use of antiarrhythmic drugs for long-term prevention of other supraventricular tachycardias (SVT) such as AV nodal reentry, reentry using an accessory pathway, and, in some centers where the expertise is available, atrial tachycardia and AFl, has taken a distant second place to catheter ablation of the critical component of the arrhythmia circuit. Antiarrhythmic drugs are now mainly used for these arrhythmias only while awaiting definitive curative ablation (7).

Pharmacological prevention of ventricular arrhythmias has also been dramatically challenged. Implantable defibrillators have been demonstrated to be superior to antiarrhythmic drugs in diminishing mortality in survivors of SD secondary to ventricular tachycardia (VT) or fibrillation in the presence of ischemic and non-ischemic CM and impaired systolic LV function (ejection fraction (EF) < 35%) (8,9). In this setting, however, antiarrhythmics can still play an adjunctive role by diminishing the number of episodes of ventricular arrhythmia or by slowing the rate of VT so as to allow painless pacing termination by the
defibrillator. Either effect should ultimately reduce the number of shocks delivered, improve the quality of life (QOL) of the patient and maximize the battery life of the defibrillator. When VT is well tolerated and LV function is relatively well-preserved (EF > 40%), long-term antiarrhythmic therapy with a drug with class III effect (sotalol or amiodarone) instead of defibrillator implantation can still be envisaged (8,9).

Antiarrhythmic drugs have so far not been successful when used as primary prevention of ventricular arrhythmias in patients at risk of SD. In general, drugs with Class I effect should not be given in either ischemic or non-ischemic CM because of increased mortality on drug (10). At best, the drugs studied to date with Class III effects have neutral effects on all-cause mortality in the post-MI patient population with congestive heart failure (CHF) (11). This latter result is in fact hypothesized to be due to a balance between an antiarrhythmic effect and a ventricular proarrhythmic effect (sotalol, dofetilide, azimilide) or the occurrence of serious organ toxicity (amiodarone). In this setting, too, the prophylactic implantation of defibrillators has proven more effective in reducing mortality compared to both placebo and amiodarone when LV systolic function is significantly reduced (EF < 35%) (12). Unfortunately, it is in the setting of CHF that the incidence of AF is highest and contributes significantly to a worse outcome. Therefore, it is necessary to demonstrate at least a neutral effect of drug on mortality in this population in order to allow its administration for atrial arrhythmias in such patients.

Clearly, there is a need for drugs which are more effective and safer in terms of proarrhythmic potential and organ toxicity than those currently available. Phase II studies should therefore compare the efficacy and safety of investigational drug X not only versus placebo but also against reference drug Y currently used in clinical practice. A pharmacokinetic profile allowing once daily dosing as well as minimal interaction with other drugs administered in such clinical settings would also be highly desirable.

II. PHASE II STUDIES TO EVALUATE THE EFFICACY AND SAFETY OF NEW ANTIARRHYTHMIC DRUGS IN ATRIAL FIBRILLATION AND ATRIAL FLUTTER

II.1. Outline of a typical development plan
Phase II studies of antiarrhythmic drugs are primarily directed at demonstrating efficacy and safety in the treatment of the most common atrial arrhythmias, AF and AFI. Various tools are available to document arrhythmia including the 12-lead electrocardiogram (ECG), Holter (24-hour) recording of 2 or 3 leads, and transtelephonic monitoring (TTM) or loop recording. The TTM loop recorder is especially useful in documenting transient episodes of arrhythmia. This can be done either automatically using preestablished criteria (rate and irregularity) for AF or by patient activation during a symptomatic episode. The device then stores in memory the preceding preprogrammed duration of recorded beats as well as an independently preprogrammed duration after activation of the storage function. The stored data can then be transmitted by telephone to a receiving station where symptoms are documented and the rhythm is interpreted.

In addition to arrhythmia documentation, the analysis of drug effect in Phase II studies also involves a detailed observation of effects on various ECG measures reflecting drug action on the sinus node (PP interval), ventricular response rate in AF or AFI, AV node conduction (PR interval) in SR, intraventricular conduction (QRS interval) and ventricular repolarization time in absolute terms (QT interval) as well as when corrected for heart rate (QTc interval). This latter measure is particularly important as an indicator of appropriate or excessive Class III effect. Other laboratory investigations are directed at modifiers of drug effect (electrolyte levels) or drug pharmacokinetics (eg., renal function tests) or at indicators of potential drug toxicity (eg., liver function tests, hematological values).

II.2. Short-term study
This trial is a multicenter, randomized, double-blind, placebo-controlled, parallel group study of the efficacy and safety of intravenous bolus administration of antiarrhythmic drug X versus reference drug Y in the acute termination of atrial fibrillation and flutter.
This study is often the first type of study done as it involves only single acute intravenous (IV) administration of drug and therefore requires only short-term monitoring of the subject. It incorporates dose ranging to search for effective plasma concentrations and doses and assesses the safety of these doses. Variations of such a study (with or without a reference drug) have been performed in the development of such Class III drugs as dofetilide (13) and ibutilide (14). It should be emphasized that the usual initial clinical approach to a patient presenting with AF/AFl should be followed, i.e., anticoagulation as recommended by current clinical guidelines and control of the ventricular response rate (6) before initiating study drug.

II.2.A. Objectives
Primary Objectives:

a. To determine the efficacy of bolus I.V. administration of investigational drug X at the highest safe dose compared to placebo and to reference drug Y in achieving conversion of AF or AFl to SR (expressed as percentage conversion within T hours of I.V. dose).

b. To determine the safety of investigational drug X at the most effective dose compared to placebo in subjects with AF or AFl.

Secondary Objectives:

a. To determine the efficacy and safety of I.V. investigational drug X at doses A, B, C and D compared to placebo in achieving conversion of AF or AFl to SR.

b. To assess the effects of doses A, B, C or D of drug X and of reference drug Y on ECG measures (sinus rate, and PR, QRS, QT and QTc intervals in responders and QRS, QT and QTc intervals and ventricular response rate in AF or AFl in non-responders).

c. To determine the pharmacokinetics of investigational drug X administered I.V. in patients with AF or AFl and its modification by concomitant cardiac or minor to moderate hepatic and renal impairment.

II.2.B Primary endpoints

a. Incidence of conversion to SR within time T of study drug administration.

b. Incidence of side effects.

II.2.C. Secondary endpoints

a. Mean time to conversion.

b. Ventricular rate in non-responders on drug treatment with respect to baseline.

II.2.D. Study Design

Subjects who meet the study inclusion/exclusion criteria will be randomized in double-blind fashion to investigational drug X at dose A, B, C or D, the reference drug Y or placebo with all to be administered as a bolus I.V. In order to assure safety of the study drug at the doses proposed, such a study could be performed in a tiered fashion:

Tier 1: Placebo versus drug X at lowest dose A versus drug X at second lowest dose B versus reference drug Y (80 subjects)

Once safety of doses A and B has been demonstrated, one could proceed to:

Tier 2: Placebo versus drug X at second highest dose C versus drug X at highest dose D versus reference drug Y (80 subjects)

With the safety of drug X at doses C and D demonstrated, the study could continue with randomization of the final 160 subjects among the six groups such that there would be an equal numbers of subjects in each group at the end of the study recruitment phase. If, however, safety is shown to be an issue at the higher doses, the study would continue with only those doses judged to be safe.
II.2.E. Planned Sample
360 subjects for 6 groups with 60 subjects per group to provide 90% power (alpha = 0.05), to
a. conclude non-inferiority of the investigational drug when there is no true treatment difference
   between the investigational drug X and the reference drug Y,
b. detect a true difference between the reference drug Y or the investigational drug X and placebo
   under the null hypothesis, and
c. have sufficient power to examine a dose-response relationship of the investigational drug at different
doses.

II.2.F. Study Population
Approximately 400 subjects with AF or AFl of duration longer than 6 hours but less than 60 days will be
randomized. A total of 360 would be expected to complete the protocol.

II.2.G. Inclusion Criteria:
   a. Either sex, between 18 and 85 years of age.
   b. AF and/or AFl lasting from 6 hours to 60 days (episodes of AF/AFl of shorter duration are likely to
      convert to SR on placebo while episodes of very long duration are less likely to respond to drug
      therapy).
   c. Anticoagulation as recommended by current therapeutic guidelines (all patients are heparinized prior
      to administration of study drug and anticoagulated for at least one month after conversion to SR;
      further, if AF/AFl has been present for > 48 hours, adequate anticoagulation (INR 2-3) must have
      been maintained for at least 3-4 weeks prior to entry into the study or intra-atrial clot must have been
      excluded by transesophageal echocardiography immediately prior to study drug administration).
   d. Calculated average (beats/12 sec x 5) resting ventricular response rate < 120 bpm achieved by
      administration of rate-control drugs (β-blockers and/or calcium channel blockers and/or digoxin).

II.2.H. Exclusion Criteria
   a. Females who are currently pregnant or breast feeding.
   b. Hemodynamic instability (CHF, hypotension, angina).
   c. Resting ventricular response rate < 70 bpm in the absence of rate-control drugs or RR interval > 3 sec;
      if a patient is on rate-control drugs, these must be adjusted so that resting heart rate is not < 70 bpm.
   d. QRS interval > 180 msec or QTc interval > 440 msec (in the presence of bundle branch block > 500
      msec) calculated on the average of 3 QT and RR intervals.
   e. Wolff-Parkinson-White syndrome which has not undergone curative ablation.
   f. History or clinical signs of thyrotoxicosis, confirmed by laboratory studies.
   g. AF or AFl from reversible non-cardiac diseases (eg., pneumonia) or from acute drug effect (eg.,
      excess caffeine, alcohol, bronchodilator therapy).
   h. History of cardiac surgery, MI, or unstable angina within the last 3 months.
   i. History of aborted SD, unexplained syncope, monomorphic or polymorphic VT.
   j. Family history of prolonged QT syndrome.
   k. Known sick sinus syndrome or atrioventricular block greater than first degree.
   l. Presence of cardiac pacemaker or defibrillator.
   m. Diastolic blood pressure (BP) > 105 mm Hg or systolic BP < 90 mm Hg.
   n. Major hematological, pulmonary (necessitating continuous oxygen therapy), hepatic or renal disease
      (eg. in the case of principal route of elimination via the kidneys of investigational drug X, calculated
      creatinine clearance (CrCl) < 20 ml/min).
   o. Plasma potassium level < 4.0 or > 5.5 mEq, or plasma magnesium level < 0.75 or > 1.25 mEq.
      Plasma potassium and magnesium may be corrected prior to study entry.
   p. Amiodarone treatment within previous 3 months.
   q. Any Class I or III antiarrhythmic agent, tricyclic or tetracyclic antidepressant, anticonvulsant, or
      phenothiazine or any other drug known to prolong the QTc within 5 half-lives before study entry.
r. Use of an experimental drug within the preceding 4 weeks.
s. Prior utilization of reference drug Y for I.V. conversion of AF/AFl.
t. History of substance abuse or dependency or ongoing psychosis.

II.2.1. Tools for assessing endpoints
Assessment Methods
Initial screening consists of a relevant medical history, physical examination, 12-lead ECG and blood sample for routine laboratory tests. An echocardiogram is obtained once the ventricular response rate has been reduced to < 120 bpm prior to study drug administration to assess LV function, valvulopathy, atrial size, and, if indicated, the presence of intra-atrial clot. A Holter recording of 3 ECG leads as well as continuous ECG monitoring (single lead) is started. Study drug is given I.V. following baseline recording for one hour.

Measure of Endpoints
ECG monitoring is continued for a maximum of 24 hours to document conversion to SR. BP and 12-lead ECGs are obtained at the start and end of the baseline period, at the time of drug administration, at 5, 10, 15, 20, 30, 60, 90, 120, 180 min, at 4, 6 and 12 hours, at the predetermined time T following study drug administration, and, finally, at the end of the study period (24 hours). A single-lead ECG may be substituted at the 5-minute intervals if obtaining a 12-lead ECG is technically impossible, but care should be taken to use a lead with an easily measurable QT interval. Preferably, the lead with the longest QT interval as determined from the 12-lead ECG should be chosen. Pharmacokinetic evaluation can be determined from venous blood samples drawn from an indwelling I.V. catheter prior to infusion of study drug (baseline) and at the times specified above.

II.2.J. Specific criteria for early withdrawal or non-utilization of subject data in analysis
a. Adverse event during drug infusion requiring stopping drug infusion; data from this subject would not be used in efficacy analysis but would still be reported in the tabulation of adverse events.
b. Protocol violation.
c. Administrative reasons.

II.2.K. Data Analysis Methods
The proportion of subjects in each treatment group achieving conversion to SR within time T can be compared using a one-way logistic model with treatment as the factor. The survival function for time to conversion can be estimated by Kaplan-Meier analysis. The following pairwise comparisons should be performed:

a. Investigational vs. reference drug: a non-inferiority test and, if positive, a superiority test of investigational drug doses A, B, C and D vs reference drug Y.

The statistical test for each of the pairwise comparisons can be made at a two-sided 0.05 statistical significance level. The predictive value of variables at baseline (e.g., duration of arrhythmia, atrial size, AF vs AFl) on the probability of conversion to SR can be assessed by multiple logistic regression analysis. Mean changes from baseline over time for mean BP, heart rate, and QRS and QTc intervals can be assessed in non-responders by analysis of covariance.

II.2.L. Comments
If safety and efficacy are demonstrated in such a study, other Phase II studies can be considered. If safety is demonstrated but there is little efficacy of I.V. drug, this does not exclude further clinical studies of drug X. Oral drug which is converted to one or more active metabolites may be effective when single dose I.V. administration is not. Further, an antiarrhythmic drug may not be useful in achieving conversion to SR but may be very effective in preventing recurrence of arrhythmia. However, if safety is clearly an issue,
especially if observed at fairly low doses and plasma levels of drug, there will be little impetus to pursuing further drug testing.

Once I.V. dosing which is both effective and safe has been established for investigational drug X, other questions best answered by physician driven Phase III studies can begin. For example, in patients failing to convert to SR, is electrical cardioversion following I.V. infusion of drug X more effective (higher percentage conversion rate or lower defibrillation threshold) than when performed in the presence of placebo or of reference drug? Further, is I.V. pre-administration of investigational drug X more effective than placebo or reference drug Y in preventing early recurrence of AF or AFl following successful cardioversion?

II.3. Long term studies

II.3.i. Conversion and maintenance sinus rhythm in patients with persistent atrial fibrillation or atrial flutter
This trial is a multicenter, randomized, double-blind, placebo-controlled parallel group study of the efficacy and safety of oral investigational antiarrhythmic drug X versus reference drug Y in converting to and maintaining SR in patients with persistent AF or AFl.

As safety has become such an issue for drugs with Class III effects, long-term Phase II studies should be initiated in a monitored setting in hospital for at least the duration of time necessary to achieve steady-state. Such an approach has been taken with both dofetilide (15) and azimilide (16) in their supraventricular Phase II programs. Also, evaluation of the efficacy of investigational drug X in the context of a long-term oral dosing protocol may now need to be done separately in patients with cardiac pathology and EF < 35 % as compared to those with better preserved EF or with normal hearts. In the former group, defibrillator implantation for SD prevention is now indicated (12,19) especially since the recent presentation of the Sudden Cardiac Death – Heart Failure Trial (12). The dual-chamber pacing capabilities of defibrillators along with various AF prevention pacing algorithms now available may significantly influence drug efficacy and safety. Further, their diagnostic abilities and memory capacity allow documentation of recurrence of atrial arrhythmia as well as of possible ventricular proarrhythmia. The following protocol therefore addresses the evaluation of investigational drug X in converting AF to and maintaining SR only in patients with normal hearts or fairly well-preserved EF (> 35%) despite cardiac pathology.

II.3.i.A. Objectives
Primary Objectives
a. To determine the efficacy of oral administration of investigational drug X at the highest tolerated dose compared to placebo and to reference drug Y in achieving conversion of persistent AF or AFl to SR within 5 half lives of the start of drug dosing (or within the time to reach maintenance plasma levels if loading doses are used).
b. To determine the efficacy of chronic oral administration of investigational drug X at the highest tolerated dose compared to placebo and to reference drug Y in maintaining SR following electrical or chemical cardioversion.
c. To determine the safety of oral administration of the investigational drug X at the most effective dose compared to placebo in subjects with persistent AF or AFl.

Secondary Objectives
a. To determine the efficacy and safety of oral investigational drug X at doses A, B and C compared to placebo in achieving pharmacological conversion of AF or AFl to SR.
b. To assess the effect of doses A, B and C of drug X and reference drug Y on ECG measures such as sinus rate, PR, QRS, QT and QTc intervals in responders or following electrical cardioversion in non-responders as well as on the ventricular response rate in non-responders.
c. To determine the range of effective plasma levels of investigational drug X in subjects with AF or AFl and its modification by concomitant cardiac or minor to moderate hepatic or renal impairment.

II.3.i.B. Primary endpoints
   a. Incidence of conversion to SR within 5 half-lives of beginning study drug or, if loading doses are used, within the time to reach maintenance plasma levels (see Study Design).
   b. Time to first recurrence of AF or AFl lasting at least 24 hours following electrical or chemical conversion to SR.
   c. Incidence of serious side effects.

II.3.i.C. Secondary endpoints
   a. Mean time to conversion in responders.
   b. Ventricular rate in non-responders with respect to baseline.

II.3.i.D. Study Design
Subjects who meet the study inclusion/exclusion criteria are admitted to telemetry in hospital for the duration of time necessary to reach the plateau effect of investigational drug X or reference drug Y whichever is longer. Randomization is done in a double-blind fashion to oral investigational drug X at doses A, B, or C, the reference drug Y or placebo. Maximum dose of investigational drug X (i.e., dose C) should not exceed that proven safe in Phase III studies of mortality. If investigational drug X is eliminated primarily by the kidneys, dosage adjustment is performed according to calculated Clcr, as follows:
   a. Clcr 40 – 60 ml/min – half of the randomized dose is given,
   b. Clcr 20 – 40 ml/min – one quarter is given.

A 12-lead ECG is obtained just before the first dose on all days while in hospital. If the QTc interval is found to increase by > 15 % over baseline, the dose is halved. If QTc interval exceeds 550 msec or increases by > 25 % over baseline, the subject is withdrawn from the study. It is important to note that in the design of such a study, reference drug Y should use the same predominant route of elimination/metabolism as the investigational drug X, as the dosage adjustments described above may penalize the reference drug Y if it does not depend on renal clearance. On the other hand, when elimination of investigational drug X does not depend on renal clearance and therefore dose adjustments are not necessary, there is a risk of significant toxicity in the case of reference drug Y when its clearance does depend on the kidneys.

In subjects who are exposed to drug for a duration of time to achieve plateau levels and have not converted spontaneously to SR, electrical cardioversion is attempted. Following conversion to SR, subjects are monitored for an additional 24 hours. Those in whom SR cannot be achieved or maintained for 24 hours following electrical or pharmacological conversion are withdrawn from the study. Subjects who are at this point in SR are discharged from hospital on the same dose of study drug as last given in hospital, or on maintenance dosing in those in which the hospitalization period was used for drug loading. Oral anticoagulation is continued for only one month in patients at low risk for cardiac emboli but throughout the course of the study in patients at moderate to high risk (6). Follow-up clinic visits are scheduled at 2, 4, 6, 8, 10, 12 weeks and then every 3 months until one of the study end points is reached: relapse to AF or AFl for at least 24 hours, documented by ECG, or maintenance of SR for one year. Twelve-month survival data (freedom from AF or AFl) is collected for all randomized subjects regardless of treatment duration.

II.3.i.E. Planned Sample
Sample size for a double-blind study can be calculated using the method of Schoenfeld (17) (proportional hazards regression model), assuming a two-sided hypothesis test of the primary endpoint at a significance level of 5 %. Total sample size will then depend on the number of groups (investigational drug X doses studied and on the presence or absence of a reference drug Y group). The distribution of time to first event documented by ECG is assumed to follow an exponential distribution. If the median time to first event is
estimated to be no more than 90 days for placebo, then a sample size of 190 subjects in each treatment group will allow a hazard ratio of 0.67 to be detected with 0.90 probability. With possibly a 10-20% drop out rate, then 210-230 subjects should be enrolled in each treatment group.

II.3.i.F. Study Population
If all the groups mentioned above are used in the study, then a minimum total sample size of 1260 subjects with AF or AFib of duration longer than 2 weeks but less than 26 weeks will be recruited. A total of about 1140 subjects would be expected to complete the protocol. If fewer doses are studied or if no appropriate reference drug Y is available for comparison, then the total population size is correspondingly reduced.

II.3.i.G. Inclusion Criteria
a. Either sex between 18 and 85 years of age.
b. AF and/or AFib lasting from 2 to 26 weeks, confirmed by ECG.
c. Anticoagulation in all patients with a therapeutic INR of 2-3 for at least 3-4 weeks prior to beginning study drug, or, if a patient is not currently anti-coagulated, heparin I.V. is administered and oral anticoagulation is begun. Study drug can be administered early on I.V. heparin if transesophageal echocardiography shows no intra-atrial thrombus.
d. Average resting ventricular response rate < 120 bpm (see Section II.2.G).

II.3.i.H. Exclusion Criteria
Same as listed in section II.2.H with the addition of the following:
a. Females who plan to become pregnant during the course of the study or, if sexually active, are not using a hormonal contraceptive as well as a vaginal spermicide.
b. Presence of severe valvulopathy, such that surgical intervention is considered a possibility within the time course of the study.
c. Ischemic or non-ischemic CM with EF < 35%.
d. Presence of other life-threatening disease with survival expected to be < 2 years.
e. Use of digoxin, unless plasma levels remain constant.
f. Absolute contraindications to anticoagulation therapy.
g. Any unresolved drug-induced organ toxicity.
h. Concomitant therapy with drugs known to affect the metabolism or elimination of investigational drug X or reference drug Y.
i. Patients who previously in the opinion of the investigator have failed an antiarrhythmic drug of the class being tested for efficacy reasons.
j. Patients who have previously participated in a study of investigational drug X or have used reference drug Y for oral conversion to and/or the subsequent maintenance of SR.
k. If Phase III studies in subjects with underlying ischemic/non-ischemic CM and but less depressed LV function (EF 35-50%) demonstrate significantly higher mortality in the presence of investigational drug X compared to placebo, such patients should also be excluded in chronic studies of efficacy in AF or AFib.

II.3.i.I. Tools for assessing endpoints
Initial screening should include a relevant medical history, complete physical examination, 12-lead ECG, chest X-ray, echocardiogram and blood sample for routine laboratory tests as well as those indicated by the known side effect profile of investigational drug X and reference drug Y. In-hospital, a 12-lead ECG is done each day before the morning dose of study drug. If electrical cardioversion is done or if chemical conversion occurs, an ECG should be obtained immediately to document SR and also 24 hours later. Blood tests before discharge from hospital should include serum chemistry, hematology, and analysis of levels of investigational drug X, reference drug Y and principal active metabolite(s), if applicable.
At follow-up visits, vital signs, cardiopulmonary specific physical examination, event symptom severity checklist (prior to ECG), 12-lead ECG, serum chemistry, hematology, blood samples for analysis of investigational drug X, reference drug Y and active metabolite(s) and assessment of concomitant medications and adverse events are done. Assessment of QOL (SF-36) and the Brignole Atrial Fibrillation Symptom Checklist can be completed on a monthly basis. As well, drug accountability and, in females of childbearing potential, a serum pregnancy test should be performed. At the final visit, either on reaching the primary endpoint or on completion of the total study duration, in addition to the procedures described for the follow-up visits, a complete physical examination, a healthcare resource utilization questionnaire and a chest X-ray are done.

A 12-lead ECG is obtained at each visit, reviewed by the investigator and sent to a central facility which generates a report including an interpretation of the patient’s rhythm and interval calculations. If the investigator and central ECG facility interpretations do not agree with respect to possible AF, AFl or SVT, the report should be sent to an Event Committee for a final decision.

Adverse Events (AE) are defined as any undesirable clinical experience during the study, whether or not related to the study drug, including an exacerbation of a preexisting condition. A serious AE is one that results in death or is life threatening, results in hospitalization or prolongation of current hospitalization, results in persistent or significant disability/incapacity, is a congenital anomaly/birth defect or is judged to be medically significant. All AEs should be judged as to their severity (mild, moderate, severe) and as to their causality (doubtful, possible, probable).

II.3.i.J Specific criteria for early withdrawal of subjects from the study
a. ECG criteria: ventricular fibrillation, polymorphic VT of any duration, sustained monomorphic VT (> 30 sec) or symptomatic non-sustained VT, incessant VT (recurrent VT < 30 sec interrupted by a few sinus beats), QTc > 525 msec, resting SR remaining < 50 bpm after conversion to SR.
b. Pregnancy
c. Noncompliance: missing any 3 scheduled visits or any 2 serum pregnancy tests, stopping contraceptive methods or taking less than 80 % of study drug.
d. Adverse events or organ toxicity such that withdrawal of the subject is recommended.
e. Voluntary withdrawal by the subject.
f. Protocol violation.
g. Administrative reasons.

II.3.i.K. Data Analysis Methods
Primary endpoints
All intent-to-treat subjects are used in the analysis of the primary efficacy endpoints. The Cochran-Mantel-Haenszel test or the Fisher’s exact test can be used to compare the percentage of randomized subjects pharmacologically converting to SR in each investigational drug X group compared to placebo. Similarly, comparison can be made between each investigational drug X group vs. reference drug Y group. Time from the start of the efficacy period to the date of the first documented AF, AFl or SVT of > 24 hours duration is measured and displayed using Kaplan-Meier estimates of the survival curves. Treatment comparisons can be made using log-rank test, differences between the time to event distributions can be quantified by the hazard ratio from proportional hazards regression, Kaplan-Meier estimates of the median time to event, and Kaplan-Meier estimates of the proportion event-free through one year. Pairwise comparisons between investigational drug X at doses A, B or C vs reference drug Y can be performed, as well as pairwise comparisons between investigational drug X at doses A, B or C vs placebo. The predictive value of variables at baseline (e.g., duration of arrhythmia, atrial size, AF vs AFI, LV systolic dysfunction) on the probability of conversion to SR or on time to recurrence of atrial arrhythmias can be assessed by multiple logistic regression analysis.
All subjects randomized to a treatment group are used in safety summaries. Incidence of AEs is compared between all groups and frequencies are tabulated by body system, treatment duration, causality and severity. Mortality rates can be calculated for each group. Treatment comparisons can be made by calculating the mortality relative rate using a Poisson model and the mortality relative risk using the Cochran-Mantel-Haenszel method (18). Efficacy and safety data can be summarized separately for subgroups defined, for example, by sex, age, race/ethnicity, baseline cardiovascular disease state, EF, smoking history, digoxin or β-blocker use and presented as point estimates and estimated standard errors. Descriptive statistics are used to calculate the clinical laboratory data and ECG intervals which can then be presented in shift tables and/or in shift plots.

Secondary endpoints
Time to conversion can be determined by the number of hours from first dose in each group and displayed graphically by means of the Kaplan-Meier (product-limit) method. Ventricular response rates in non-converters at first ECG before study drug administration and prior to electrical cardioversion can be presented in the different groups with descriptive statistics. Change and percentage change can be compared between placebo, investigational drug X groups and reference drug Y group by analysis of variance and Bonferroni t-test. QOL comparisons can be made using analysis of variance or if assumptions of normality and equal variances are not valid, a non-parametric analysis of covariance can be used. Symptom frequency load during the first recurrence can be assessed by constructing contingency tables of number of treatment groups versus number of pre-specified symptoms reported during the first recurrence of atrial arrhythmia and analyzed with the appropriate chi-square test.

II.3.i.L. Comments
Demonstration of efficacy and safety in such a study suggests the possibility of beginning drug on an outpatient basis. However, if the dose must be modified during the first few days either because of excessive effects on the ECG or because of significant proarrhythmic effects, further Phase II studies will require the same initiation protocol, and, once brought to market, initiation in a monitored hospital setting will be mandatory. This would significantly limit this drug’s use appropriately to specialists trained in its administration and, perhaps inappropriately, for those for whom financial restrictions and hospital bed availability are not an issue.

Further Phase III studies could determine if defibrillation thresholds for AF are decreased by oral pretreatment with investigational drug X versus placebo or reference drug Y, or in patients with pacemakers or pacemaker/defibrillators with programming capabilities for pacing prevention of AF and AF1, it could be determined if such programs are more effective in preventing recurrence of these arrhythmias in the presence of drug X versus placebo or reference drug Y.

II.3.ii. Prophylactic treatment of paroxysmal atrial fibrillation or atrial flutter
The trial is a multicenter, randomized, double-blind, placebo-controlled, parallel-group clinical study to assess the efficacy and safety of the oral investigational antiarrhythmic drug X versus reference drug Y in the prophylactic treatment of paroxysmal atrial fibrillation and atrial flutter.

With the accumulation of data from studies on the acute conversion of AF and AF1 with I.V. or oral investigational drug X and the prevention of recurrence of arrhythmia following pharmacological or electrical cardioversion, it is usually possible to identify a dose range which is effective and tolerated. The number of investigational drug groups can now be limited to one or two doses for evaluating efficacy and safety when used for other indications, such as the prevention of paroxysmal AF (PAF). Accumulating data from mortality studies in subjects with moderate systolic LV dysfunction in ischemic or non-ischemic CM will determine whether such a study should be limited to subjects with PAF or PAF1 and normal hearts or whether two strata can be studied as described below. If so, as in section II.2, such a chronic study should
exclude patients with ischemic or non-ischemic CM and LVEF < 35% for whom defibrillator implantation is indicated.

II.3.ii.A. Objectives
Primary Objectives
a. To assess the efficacy of oral investigational drug X at doses A and B compared to placebo and reference drug Y in prolonging the arrhythmia-free period, i.e., time from first dose of study medication to first ECG-documented symptomatic AF, AFI or SVT.
b. This primary objective can be assessed separately in two strata:
   • subjects with ischemic heart disease (IHD) and/or CHF (EF >35%)
   • subjects with neither of the above

Secondary Objectives
The following secondary objectives can be determined according to stratification between presence or absence of CHF and IHD.

a. To assess the efficacy of investigational drug X at doses A and B versus placebo and versus reference drug Y in reducing the total number of symptoms reported during the first symptomatic event.
b. To assess the efficacy of investigational drug X at doses A and B versus placebo and versus reference drug Y in reducing the frequency of symptomatic events.
c. To assess the efficacy of investigational drug X at doses A and B versus placebo and versus reference drug Y in reducing the total supraventricular arrhythmia burden in patients (i.e., frequency, duration and severity).
d. To assess the effect of investigational drug X at doses A and B versus placebo and versus reference drug Y on patient QOL.
e. To assess the number of days in-hospital and number of emergency room visits due to symptomatic PAF, PAFI or PSVT (following initial discharge for subjects hospitalized for drug loading or initiation) between groups receiving investigational drug X at doses A and B versus placebo and versus reference drug Y.
f. To assess the efficacy of investigational drug X at doses A and B versus placebo versus reference drug Y in reducing the number of asymptomatic events.

And, in both strata combined:
g. To assess the safety of investigational drug X at doses A and B versus placebo and versus reference drug Y according to incidence and severity of AEs, ECG changes, and new laboratory and chest X-ray abnormalities.

II.3.ii.B. Study Endpoint
Time to a SR containing day after the second confirmed occurrence of symptomatic PAF, PAFI or PSVT.

II.3.ii.C. Study Design
Subjects who meet the study screening inclusion/exclusion criteria are instructed in the use of a TTM. SR at the beginning of the screening period is documented during a training call. Over the following month, subjects transmit a recording obtained whenever they have symptoms suggestive of an arrhythmia. Once symptomatic AF or AFI is documented, subjects transmit again when SR returns. Randomization then occurs between investigational drug X at dose A or B, placebo or reference drug Y. If no spontaneously terminating episode of AF or AFI is documented during a one-month screening phase off antiarrhythmic drugs, the subject is withdrawn from the study. Study drug is initiated only once other Class I or III drugs are stopped for at least 5 half-lives.
Initiation of study drug using in-patient hospitalization, continuous monitoring and a 12-lead ECG prior to each morning dose versus close out-patient monitoring and daily transmission of a rhythm strip prior to morning dose will depend on the incidence of proarrhythmia observed in prior trials with investigational drug X in the absence and the presence of known IHD or CHF as well as on the clinical experience with reference drug Y. In the case of significant renal clearance of drug as well as risk of torsade de pointes, in-hospital initiation of study drug using the dosing design described in section II.2.D should be used. When clearance is primarily by hepatic metabolism and there is little risk of torsade de pointes, out-patient initiation of study drug can be considered, at least in cases of PAF or PAFI in normal hearts.

During the maintenance period of study drug administration, a TTM recording should be transmitted whenever the subject experiences symptoms suggestive of arrhythmia. An Event Symptom Severity Checklist should be obtained by the central TTM facility before transmission. If the subject is in a location where an ECG can be obtained, this checklist should be administered by health care personnel prior to recording the rhythm. All tracings should be sent to the central facility. Documentation of return to SR can be made by TTM, ECG or telemetry. If return to SR does not occur spontaneously, then Class II and/or Class IV drugs and/or digoxin can be administered to slow the ventricular response rate to < 100 bpm, and, if arrhythmia persists, electrical cardioversion can be done. Neither Class I nor Class III drugs should be used. Subjects then continue on the same study drug assigned until documentation of return to SR after the second confirmed symptomatic atrial arrhythmia occurs or until end of study at 12 months.

II.3.ii.D. Planned Sample
Sample size is calculated as in section II.2.E using Schoenfeld’s method, assuming two-sided tests, and a time to first event that follows an exponential distribution with a median of 90 days in the placebo group. Since the assigned alpha applies only to the time to first symptomatic event, Hochberg’s procedure is applied to the sequence of the 2 tests (17). Allowing for drop-outs, approximately 220 subjects are projected for each group. This would allow an investigational drug to placebo hazard ratio of 0.67 to be detected with 0.90 probability at an alpha of 0.045.

II.3.ii.E. Study Population
If the investigational drug is administered at doses A and B, and both placebo and reference drug Y are used, then a study sample size of 880 subjects with PAF or PAFI are required in each stratum after successful screening. It should be pointed out that the logistics of screening a large patient population to document a symptomatic arrhythmia episode within a one-month period may significantly dampen enthusiasm for a multi-group trial. To be practical and finish the study recruitment in a timely manner, only the most promising dose of investigational drug X may be compared to placebo. Comparison to a drug already on the market may have to wait for positive results of a simpler investigational drug X versus placebo trial.

II.3.ii.F. Inclusion Criteria
Screening Period
a. Either sex between 18 and 85 years of age.
b. Symptomatic AF within the previous 6 months, documented by ECG, TTM, or rhythm strip but not by Holter monitor.
c. Symptoms severe enough to significantly interfere with the subject’s usual activities.
d. SR at the time of starting the screening period.

For Randomization
a. Within 30 days from starting screening, at least 1 episode of symptomatic AF or AFI documented by TTM or ECG (if no episode occurs on antiarrhythmic drugs, at least 1 episode during a second 30-day screening period after discontinuing these medications).
b. SR immediately prior to starting study drug.
II.3.ii.G. Exclusion Criteria

Screening Period

a. Women currently pregnant or breast-feeding, or who plan to become pregnant during the course of the study or are unwilling to use hormone contraceptives with a vaginal spermicide; women postmenopausal for less than one year or not surgically sterile (tubal ligation at least 3 months prior to screening) unless they have been using hormonal contraceptives for at least 3 months prior to entry into the screening phase.

b. AF or AFL due to acute electrolyte disturbance, hyperthyroidism, pericarditis, or any other acute reversible illness.

c. Electrical cardioversion within 60 days prior to screening.

d. Syncope, angina pectoris or pulmonary oedema that is precipitated by attacks of arrhythmia (ventricular or supraventricular).

(e) Polymorphic VT of any duration, sustained (> 30 sec) monomorphic VT, aborted SD or undiagnosed syncope.

f. Wolff-Parkinson-White syndrome that has not undergone curative ablation.

g. QRS interval > 180 msec or QT interval > 440 msec (in the presence of bundle branch block > 500 msec) or a family history of prolonged QT syndrome.

h. Presence of an implanted pacemaker or defibrillator.

i. History of Class IV New York Heart Association CHF or current acute decompensated CHF. Class I – III patients must have been stable for at least one month.

j. Ischemic/non-ischemic CM with EF < 35%.

k. Cardiac surgery, MI, or unstable angina within the last 3 months.

l. Symptomatic severe valvular disease for which surgery is considered within the duration of the study.

m. Stroke or a reversible ischemic neurologic deficit within the last 3 months.

n. Diastolic BP > 105 mm Hg (acceptable value must be present for at least one month on antihypertensive therapy) or systolic BP < 90 mm Hg.

o. Major hematological, pulmonary (requiring continuous oxygen therapy), hepatic or renal disease (calculated ClCr < 20 ml/min).

p. Unresolved drug-induced organ toxicity.

q. Substance abuse or dependency or ongoing psychosis.

r. Previous failure of efficacy of the class of antiarrhythmic drug being tested.

s. Prior participation in a randomized trial of investigational drug X.

t. Prior exposure to reference drug Y.

u. Any investigational drug use in the 30 days before screening.

For Randomization

a. Electrical cardioversion during the screening period.

b. Resting SR below 50 bpm.

c. QTc interval > 440 msec, calculated as the average of 3 QT and RR intervals recorded during SR.

d. Class I or Class III drugs within 5 half-lives or amiodarone within one month prior to receiving the first dose of study drug or during the study.

e. Use of drugs which prolong the QTc interval when the investigational drug X or reference drug Y is a Class III drug and of any drug that may potentiate a known serious side effect of investigational drug X or reference drug Y for at least 5 half-lives prior to receiving the first dose of study drug or any time during the study.

f. Potassium < 4 mEq/L or > 5.5 mEq/L or magnesium below the lower limit of normal; both of these electrolyte levels may be corrected prior to beginning study drug.

g. Hepatic dysfunction (ALT, AST > 2 x normal), severe renal dysfunction (calculated ClCr < 20 ml/min).
II.3.ii.H. Tools for assessing endpoints
A baseline visit for randomized subjects is performed within 7 days prior to first dose of study drug. This includes a medical history, complete physical examination, chest X-ray, hematology, blood chemistry (electrolytes including magnesium, renal and hepatic function tests and others as suggested by the side effect profile of both investigational drug X and reference drug Y), blank sample for drug and metabolite plasma levels and serum pregnancy test for women of child-bearing potential. An echocardiogram should have been done to determine eligibility for the study prior to screening. Potassium and magnesium levels must at all times be maintained in the normal range. Immediately prior to the first dose of study drug, vital signs are taken, an event severity checklist is done and then a 12-lead ECG is taken to confirm SR and QTc < 440 msec and serum chemistry is repeated stat if baseline chemistry is not done on day 1 of dosing. Finally, QOL is assessed by SF-36 and Brignole Atrial Fibrillation Symptom Checklist.

Clinic visits and TTM are scheduled often to document asymptomatic/minimally symptomatic recurrence of atrial arrhythmia. TTM transmission can be done weekly if no clinic visit is scheduled for that week with Event Symptom Severity Checklist done before transmission. Clinic visits can be scheduled at the end of week 2, 4, 8, 12, 16, 24, 36, and 52. At each visit the following are recorded: vital signs, Event Symptom Severity Checklist before a 12-lead ECG, blood chemistry, hematology, blood levels of investigational drug X, reference drug Y and possible metabolites, assessment of adverse events and verification of concomitant medications. Also, drug accountability is done at the beginning of the maintenance phase and both drug accountability and QOL assessment are done at weeks 4, 8, 12 and 24 and 52. A pregnancy test is done at monthly intervals in women of child-bearing potential following the baseline test until the end of the patient’s participation in the study. At the last visit (week 52 or on early completion or withdrawal) a complete physical examination is performed and a chest X-ray should be obtained.

Post-treatment follow-up is done at day 7 and day 30 following study termination. The same data is collected as above as well as a pregnancy test if indicated at day 30 post-treatment. During this 30-day period subjects continue to transmit recordings during symptoms suggestive of arrhythmia. The TTM is returned only at the end of this 30 day period.

Twelve-lead ECGs obtained at each scheduled or unscheduled visit and any TTM or telemetry recordings are reviewed by the investigator. All the above are sent to a central facility which will generate a report of rhythm diagnosis and interval measurements. If there is disagreement with respect to the rhythm diagnosis, the tracing should be submitted to an Event Committee. QOL questionnaires will be completed at home prior to the appropriate visit. Safety (AEs) will be assessed as described in section II.3.K.

II.3.ii.I. Data Analysis Methods
Primary endpoints
The log-rank test can be used to compare the distribution of the time to first symptomatic event (from day of randomization) between patients in the various groups. Hochberg’s procedure can be applied to the sequence of the 2 tests (17). Differences between the time to event distributions can also be quantified by the hazard ratio from a proportional hazards regression, Kaplan-Meier estimates of the median time to event, and Kaplan-Meier estimates of the proportion event-free through 6 months or 12 months. The presence or absence of Class II/III CHF can be used as a stratification variable for the log-rank test and proportional hazards regression in the CHF/IHD stratum.

Secondary endpoints
The number of 6 pre-specified symptoms reported during the first confirmed event (Event Symptom Severity Checklist) are totaled. The total count can be compared between treatment groups, using a chi-square test with rank scores. Also, the percentage of subjects with each specific symptom can be compared using a chi-square test. Symptomatic event rates can be compared between treatment groups and analyzed using a Poisson regression model (18). ECG data can be used to assess subject arrhythmia status over time.
and this latter can be combined with symptom count and severity data to classify each subject into an ‘SVA burden state’. The proportion of subjects in each state or combination of states can be compared between treatment groups using a logistic regression or hazard regression model.

The Medical Outcome Study Short Form (SF-36) Health Survey consists of 36 questions which assesses 8 areas one of which is physical functioning. This subscale can be normalized to a 0 – 100% scale from worst to best functioning. Descriptive statistics (n, mean, median, minimum, maximum and standard errors) can be used to summarize the data and analyses of variance can be used to test for treatment differences in physical functioning. Resource utilization can be assessed by evaluating the number of days in-hospital after in-hospital initiation of study drug or the number of emergency room visits due to atrial arrhythmias normalized by the number of days at risk. This can be modeled using Poisson regression and the relative risk and standard error in investigational drug X group versus placebo or versus reference drug Y can be estimated.

Other analyses can compare the rate of asymptomatic arrhythmias between treatment groups using a Poisson regression model. Other SF-36 QOL measures (eg., general health, vitality, social functioning and mental health) can also be assessed as described above. The Brignole Atrial Fibrillation Symptom Checklist quantifies the presence of 5 symptoms which are disease specific (palpitations, shortness of breath at rest, shortness of breath during physical activity, fatigue during mild physical activity and fatigue at rest) during the past 4 weeks by means of a score scale (0- absent to 10- present). This too can be summarized with descriptive statistics or frequency tables at the scheduled time points. Analysis of variance can be used to test for treatment differences using factors such as baseline value, sex or baseline EF as covariates.

The incidence of AEs can be summarized using tables. Descriptive statistics can be used to summarize laboratory data as well as ECG values (QT, QTc, PR, QRS and heart rate during sinus rhythm or during atrial arrhythmia) by visit, including absolute values, change and percentage change from baseline. The latter can be presented in shift tables or plots. Reason for drug discontinuation should be presented in frequency tables. Compliance (number of tablets taken divided by number expected to be taken) can be summarized with descriptive statistics.

II.3.ii.J. Specific criteria for early withdrawal or non-utilization of subject data in analysis

a. ECG criteria: ventricular fibrillation, sustained monomorphic VT (> 30 sec), incessant VT (recurrent VT with episodes lasting < 30 sec and interrupted by a few sinus beats), polymorphic VT or a QTc > 525 msec on 12-lead ECG or rhythm strip. If a patient is withdrawn from the study for ECG criteria, a blood sample should be obtained for plasma level of investigational drug X and reference drug Y as well as any possible active metabolites and for electrolytes (including Mg++).
b. Pregnancy.
c. Noncompliance: missing any 3 or 2 consecutive clinic visits, a total of 3 or any 2 consecutive TTM transmissions, or 2 serum pregnancy tests; stopping hormonal contraceptives or vaginal spermicide; use of any excluded concomitant medication.
d. Adverse events or organ toxicity that require subject withdrawal from the study.
e. Protocol violation.
f. Administrative reasons.

II.3.ii.K. Comments

The above Phase II studies represent the core of investigational drug development for the indication of AF or AFL. They provide an excellent evaluation of drug efficacy and safety of both I.V. and oral forms at various doses compared to placebo. Although to date not often done within the same study, incorporating the most effective reference drug Y also provides useful information. It is such a comparison which in
fact determines whether investigational drug X is eventually marketed just as a useful alternative to presently used drugs or whether it can in fact be considered the drug of first choice.

III. PHASE III STUDIES IN PATIENTS AT RISK OF SUDDEN DEATH DUE TO VENTRICULAR ARRHYTHMIAS

The Cardiac Arrhythmia Suppression Trial (1,2), a Phase III study of safety and efficacy of flecainide and encainide in a large patient population at risk of SD in the chronic phase of MI, was carried out well after the introduction of these two drugs to market. Since the finding in this study that mortality was increased on drug, such Phase III testing has been more recently done early and, in fact, prior to obtaining regulatory approval. Indeed, observations made during such a study often can help in designing safer Phase II studies. For example, the most recently published mortality trial compared azimilide to placebo in patients with recent MI (6-42 days) and who were at moderate to high risk for SD as predicted by poor LV systolic function (EF < 35%) with or without low heart rate variability (20). No difference in overall mortality was observed on drug. Although azimilide would therefore not be indicated for prevention of SD, such a study does confirm the safety of its use for atrial arrhythmias in this population and permits inclusion of such patients in clinical trials of drug for supraventricular arrhythmias.

It is unlikely, however, that such placebo-controlled mortality studies in moderate to high risk populations will continue to be performed. The recent publication of MADIT II (19) and the presentation of the SCD-HeFT results (12) have shown the superiority of implantable defibrillators versus placebo in patients with poor LV function and IHD as well as versus placebo and amiodarone in non-ischemic CM (12). The indication now exists, therefore, for implantation of a defibrillator in such patients. As a consequence, long-term mortality trials of oral investigational drug X versus placebo are, in our opinion, no longer ethical in this population. Instead, the focus will likely be shifted to that group of patients with ischemic or non-ischemic CM and with only moderately or minimally depressed LV function (EF 36-50%). These patients, while less at risk for SD from ventricular arrhythmias as a result of the underlying cardiac pathology, may however be at risk for greater mortality on investigational drug X if there is any proarrhythmic potential of the new drug. Such a suggested study is presented.

III.1. Outline of a typical development plan
The study should be a double-blind, placebo-controlled, parallel design to determine the effect of two doses of orally administered investigational drug X versus placebo on survival in ischemic or non-ischemic cardiomyopathy and low to moderate risk of sudden death.

III.2. Long-term studies

III.2.A. Objectives
Primary Objective
To evaluate the effects of investigational drug X at dose A versus placebo or dose B versus placebo on all-cause mortality based on intention-to-treat in subjects at low to moderate risk of SD, i.e., with either ischemic or non-ischemic CM and LV EF 36-50 %.

Secondary Objectives
To evaluate the effects of investigational drug X at doses A and B combined versus placebo on all-cause mortality, based on intention-to-treat analysis.

Tertiary Objectives
These analyses will be based on ‘on-treatment’ observations, i.e up to 30 days following discontinuation of the study drug.
a. To evaluate the effects of investigational drug X at dose A versus placebo or dose B versus placebo on all-cause mortality.
b. To evaluate the effects of investigational drug X at doses A and B combined on arrhythmic, cardiac, and non-cardiac mortality.
c. To determine the effect of beta blocker and angiotensin converting enzyme inhibitor use on all-cause mortality.

III.2.B. Study Endpoint
The study endpoint is death due to any cause.

III.2.C. Study Design
Patients who meet the study inclusion/exclusion criteria are enrolled in the study and first undergo a 24-hour Holter to determine heart rate variability (HRV) which is analyzed at a central facility. The baseline visit (within 48 prior to initiation of study drug) also includes a complete physical examination, review of concomitant medications, 12-lead ECG, stat potassium and pregnancy test (if applicable) and chemistry and hematology measurements.

The study subjects are then equally randomized among the treatment groups. Short-term hospitalisation during initiation of treatment to achieve plateau drug levels and effect on ECG intervals will depend on the need for continuous monitoring because of risk of early proarrhythmia as suggested by early I.V. or oral single-dose Phase II studies. Dosing for investigational drug X whose elimination is highly dependent on renal function will be adjusted as described in section II.3.D. Further adjustment of drug dose in the case of a Class III drug will also be influenced by the effect on QTc interval as described in the same section.

Total study duration is two years. Follow-up visits can occur at 2 weeks, and 1, 4, 8, 12, 16, 20 and 24 months and 1 month following the end of the study or 4 weeks following withdrawal from the study. The same evaluations as at baseline, except for the Holter recording, are done at each of these visits. As well, drug compliance evaluation can be performed at months 1, 4, 8, 12, 16, 20 and 24. Serum pregnancy test should be done monthly in patients of childbearing potential. Patients who withdraw early from the study should be followed by telephone contact until a time corresponding to two years since study drug initiation (for assessment of outcome on an intention-to-treat basis).

III.2.D. Planned Sample
If it is assumed that the all-cause mortality in such low to moderate risk patients is 5-8% over two years, and investigational drug X will reduce this rate by at least 45%, then using the method of Schoenfeld (17), to ensure 90% power at a significance level of 0.04, at least 4000 subjects must be recruited to show significantly reduced mortality on investigational drug X. If, however, the purpose of the study is to only demonstrate safety of drug in this population, then a hypothesis of non-inferiority would require a smaller population sample size.

III.2.E. Study Population
A total of approximately 4000 subjects with ischemic or non-ischemic CM and LV EF of 36 - 50% will be enrolled.

III.2.F. Inclusion Criteria
a. Ischemic or non-ischemic CM.
b. LV EF (36-50%).
c. Male or female between the ages of 18 and 75 years.
d. If female, post-menopausal for more than one year, surgically sterile (at least 3 months post tubal ligation) or on oral contraceptives for at least 3 months prior to study entry which they will continue to use in addition to a vaginal spermicide during the study.
III.2.G. Exclusion Criteria

a. If female, currently pregnant or breast feeding, or plan to become pregnant during the course of the study.
b. History of torsade de pointes or any form of polymorphic VT.
c. History of sustained (> 30 sec) monomorphic VT.
d. Syncope or aborted SD.
e. Resting heart rate below 50 bpm.
f. Second-degree (Mobitz II) or third degree AV block without a permanent pacemaker.
g. Implantable cardiac defibrillator.
h. QTc interval measuring 450 msec or greater at enrollment or a family history of long QT.
i. Wolff-Parkinson-White syndrome not having undergone curative ablation.
j. Decompensated CHF at the time of enrollment.
k. Unstable angina pectoris.
l. Angioplasty or coronary artery bypass grafting within one month prior to study entry.
m. Severe symptomatic valvular disease for which surgery is considered within the time course of the study.
n. Stroke with significant neurological deficit.
o. Uncontrolled hypertension (systolic BP > 170 mmHg or diastolic BP > 100 mmHg) at enrollment. BP must be controlled adequately for one month prior to study entry.
p. Known concurrent illness likely to affect survival within 2 years of study entry.
q. History of chronic liver disease or significant kidney disease (CrCl < 20 ml/min).
r. History of unresolved organ toxicity secondary to drug use.
s. Prior to randomization, baseline potassium < 4 mEq/L or > 5.5 mEq/L. Potassium level can be corrected and must be maintained within the acceptable range during the study.
t. Amiodarone use within one month prior to enrollment; current use of Class I or III antiarrhythmic drugs or within 5 half-lives prior to beginning study drug.
u. Use of drugs which prolong QTc interval during or within 5 half-lives prior to study.
w. Use of any non-approved investigational drug within 30 days prior to enrollment.
x. Previous participation in a trial of investigational drug X.
y. Known alcohol abuse or illicit drug use or current, diagnosed psychosis.
z. Unwillingness or inability to give written informed consent.

III.2.H. Tools for assessing endpoints

Measure of LV EF can be performed by nuclear isotope or invasive contrast ventriculography, or by 2D-echocardiography. Low HRV on 24-hour Holter is assessed by standard methods and defined as less than 20 U. A 12-lead ECG to assess study drug effects on QTc interval is performed prior to and daily during study drug initiation if indicated with in-hospital monitoring for dose adjustment, and at each follow-up visit.

All-cause mortality is documented, but is further subclassified according to the following definitions:

a. Cardiac mortality: all deaths except those due to a demonstrated non-cardiac cause.
b. Non-cardiac mortality: death due to demonstrated non-cardiac cause including vascular mortality (stroke, embolism, ruptured aneurysm).
c. Arrhythmic mortality: - death within 1 hour on the onset of new symptoms in the absence of severe LV dysfunction or shock; unwitnessed death in an apparently stable patient; unresuscitated ventricular fibrillation; sudden or non-sudden cardiac death with documented or suspected arrhythmia; unwitnessed, sudden, presumed cardiac death.

Adverse events are defined as in section II.3.K.
III.2.I. Data Analysis Methods
The primary efficacy analysis is performed on all-cause mortality in all randomized patients. Kaplan-Meier curves are used to estimate survival for each treatment group. Median time-to-event, with 95% confidence intervals, are determined for each group. More complex statistical analyses can also be performed for each active to placebo comparison. Secondary efficacy analyses based on intention-to-treat observations are performed as for the primary efficacy analysis. The same statistical methods are used for the on-treatment efficacy analyses but data from patients who withdraw early from the study is gathered only until one month following discontinuation of study drug. Interim efficacy analyses can be conducted at pre-determined times during the course of the study and should be based on all-cause mortality. Both, statistically significant decrease or increase in mortality on investigational drug X will be assessed but the study should only be prematurely terminated for a statistically significant increase in all-cause mortality on investigational drug X.

Safety data can be presented using graphs and tables of summary statistics by treatment group. Frequency of AEs, AEs per unit time of study drug exposure are shown. Laboratory data (chemistry, hematology and ECG measurements) are shown as box plots over time. Further, ECG data can be summarized as actual values, change from baseline and percent change from baseline. Efficacy and safety data can be summarized separately for such subgroups as sex, age, ischemic versus non-ischemic CM, Class I versus Class II/III CHF or utilization of β-blockers and/or angiotensin converting enzyme inhibitors.

Interim analyses should be done at predetermined intervals by a statistician not associated with the study and are based on the primary end-point of all-cause mortality. Both, statistically significant improvement or decrease in mortality on investigational drug X should be noted, but recommendation to prematurely terminate the study should only be made if decreased survival is observed on drug X.

III.2.J. Specific criteria for early withdrawal or non-utilization of subject data in analysis
Early withdrawal of subjects from the study is based on the same criteria as in section II.3.J however these data are still used in the intention-to-treat efficacy analyses. Subject data only is not used in the case of protocol violation or for administrative reasons.

III.2.K. Comments
The demonstration of increased mortality in the presence of investigational drug X in patients with underlying cardiac pathology and mild to moderate LV dysfunction would essentially limit the use of the new drug to patients with normal hearts experiencing supraventricular arrhythmias. Even in this setting, long-term use should be frequently reevaluated as the patient ages and develops cardiac disease. If there is a need for in-hospital initiation of drug because of significant incidence of proarrhythmia even in the absence of cardiac pathology, there would be little hope that the new antiarrhythmic drug would be accepted, either by the regulatory boards or by the medical community. In contrast, if decreased or an unaffected mortality rate is demonstrated, this would permit drug use for atrial arrhythmias in patients with moderate LV dysfunction (EF 36-50%).

IV. CONCLUSIONS
Clinical trials of potentially useful antiarrhythmic drugs are at the present time both complex and costly. For the time being, the potential market of patients with AF justifies the effort and expense to complete such studies in order to obtain regulatory board approval. This situation may improve in the future if other classes of drug action are discovered which do not cause proarrhythmia as do the Class I and III compounds. In the meantime, close attention is being paid by the pharmaceutical industry to the ongoing development of and indications for implantable defibrillators in the prevention of SD due to ventricular arrhythmias, and in the case of atrial arrhythmias, to the enormous strides being made in catheter-based ablation techniques.
V. REFERENCES


12. The Sudden Cardiac Death – Heart Failure Trial (SCD-HeFT). Late Breaking Clinical Trial Results. Annual Meeting of the American College of Cardiology, New Orleans, 2004.


Pharmacological Research in Neurologic Disorders

Chapter 16. Epilepsies and Convulsive Disorders

Chapter 17. Headache

Chapter 18. Alzheimer’s Disease and Other Dementias

Chapter 19. Parkinson Disease and Other Extrapyramidal Disorders

Chapter 20. Multiple Sclerosis and Other Demyelinating Diseases
Chapter 16. Epilepsies and Convulsive Disorders

Santiago Arroyo, M.D., Ph.D. 1
David W. Chadwick, M.D. 2
Jacqueline French, M.D. 3
Richard Mattson, M.D. 4
Emilio Perucca, M.D., Ph.D. 5

1 Medical College of Wisconsin
Froedtert Hospital
9200 West Wisconsin Avenue
Milwaukee, WI 53226
USA
Tel +1-414 805 5337
Fax +1-414 259 0469
Email sarroyo@mcw.edu

2 Walton Centre for Neurology and Neurosurgery
Liverpool
UK

3 Department of Neurology
University of Pennsylvania School of Medicine
Philadelphia, PA
USA

4 Department of Neurology
Yale University School of Medicine
New Haven, CT
USA

5 Clinical Pharmacology Unit
Department of Internal Medicine and Therapeutics
University of Pavia
Pavia
ITALY

Corresponding Author:
Santiago Arroyo, M.D., Ph.D.
I. INTRODUCTORY REMARKS

With an overall prevalence in the order of 0.5 to 1%, epilepsy is a very frequent serious neurological disorder, and almost invariably requires long-term pharmacological management. Over the past 15 years, several new antiepileptic drugs (AEDs), including felbamate, fosphenytoin, gabapentin, lamotrigine, levetiracetam, oxcarbazepine, tiagabine, topiramate, vigabatrin and zonisamide, have appeared in the market with the objective of improving efficacy, tolerability, and ease of use compared with the classical AEDs (carbamazepine, valproic acid, phenytoin, phenobarbital, primidone, ethosuximide and benzodiazepines). These objectives, however, have not been completely fulfilled. Most importantly, the introduction of the new AEDs has had little impact towards the goal of achieving seizure freedom in the many patients (about one third of a typical unselected epilepsy population) who remain refractory to conventional medications.

There are many areas where clinical research is needed. The comparative efficacy and tolerability of the newer AEDs is not yet clearly defined and there is a paucity of trials comparing head-to-head these agents. Even direct comparisons between newer and older AEDs have been relatively scarce and limited to selected epilepsy syndromes (mostly patients with partial seizures with or without secondary generalization, and patients with primarily generalized tonic-clonic seizures). Moreover, most of these trials had questionable designs in terms of low sample size, suboptimal dosing regimens (often favoring the sponsor’s product), and limited duration of follow-up. Studies in generalized epilepsy syndromes and in special populations such as infants, children and the elderly have been particularly scarce.

Despite the relative abundance of different AEDs, clinicians feel that there are still unmet needs in the pharmacological treatment of epilepsy. First, over a third of the patients with epilepsy are not controlled with the current options. Second, up to a quarter of the patients exposed to a first AED will have adverse effects severe enough to require the drug’s withdrawal, and many more will suffer from chronic adverse effects limiting their quality of life. Thirdly, several epilepsy syndromes remain resistant to standard therapies. Examples include the Lennox-Gastaut syndrome and mesial temporal sclerosis. Newer drugs with improved efficacy and tolerability profiles are surely needed.

Despite the relatively “crowded” market, pharmaceutical companies should be interested in further development of newer antiepileptic compounds. Not only is epilepsy frequent and will increase further in frequency as the elderly population enlarges, but also many AEDs have found additional indications for other CNS disorders (e.g., migraine prophylaxis, neuropathic pain, anxiety, and bipolar disorder) that amplify the rewards of this line of research.

As seizures are potentially dangerous events, placebo as the only treatment has generally not been accepted in AED research. Therefore, AEDs are investigated initially as adjunctive therapy in patients with partial seizures refractory to available medications. However, due to the fact that newly diagnosed epilepsy is treated with a single AED, a monotherapy license is highly desirable and ultimately monotherapy trials also need to be conducted. These are usually done prior to licensing (conversion to monotherapy) or after a license for adjunctive therapy use has been obtained (monotherapy trials in patients with newly diagnosed partial and/or generalized seizures).

Clinical trials in epilepsy are largely geared towards registration in the US by the Food and Drug Administration (FDA) and in Europe by the European Medicines Evaluation Agency (EMEA). Unfortunately, registration trials leave many questions unanswered. For example, most studies are performed in either newly diagnosed patients or those with severe, refractory epilepsy, but few if any are performed in patients with less severe but established epilepsy. Certain common epilepsy syndromes, such as absence and juvenile myoclonic epilepsy are difficult to study in a randomized controlled way, and as a result there is very little useful data on the efficacy of the new AEDs in these populations. Also, certain populations, such
as the developmentally disabled, or those with psychiatric and medical comorbidities are often excluded from randomized trials, again leading to a paucity of data about safety concerns, if any, in these groups.

II. PHASE II STUDIES FOR REGISTRATION OF NEW ANTIEPILEPTIC DRUGS

II.1. Outline of a typical development plan

This phase usually begins with open label exploratory studies to assess titration rates, maximally tolerated dosages, pharmacokinetics and drug interactions. Results of these studies are critical for the appropriate design of pivotal studies. Determining the influence of other antiepileptic drugs on the study drug is particularly important, and should be investigated using well designed interaction studies. Many standard AEDs will either induce or inhibit the metabolism of a study drug, if it is cleared via hepatic microsomal enzymes. Failure to discover such an interaction in early phase II could lead to substantial underdosing or overdosing in the pivotal trial. Similarly, if the study drug inhibits the metabolism of a commonly used background AED such as phenytoin or carbamazepine, interpretation of efficacy outcome in a randomized trial could be severely confounded. Subsequently, the drug candidate is assessed against placebo as adjunctive treatment in patients refractory to standard therapies. As a rule, these studies are carried out in patients with partial seizures, because these patients are easier to recruit. Studies seek to enroll patients with a relatively high frequency of seizures, to permit evaluation of clinical response over a relatively short time scale.

Two types of randomized controlled designs have been used: cross-over and parallel. Due to difficulties in carrying out cross-over trials and problems with interpretation of results, a parallel-group design is preferred.

The typical pivotal trial will utilize a multicenter, double-blind placebo-controlled randomized, adjunctive therapy, parallel group design. At least 2 or 3 dose levels should be explored, preferably within the same trial. Early proof-of-concept monotherapy studies (presurgical design) may occasionally be included as part of the phase II development program (see section 3). Patients included in short-term studies should be allowed to enter long-term open-label follow-up.

II.2. Short-term studies

Adjunctive therapy trial in patients with refractory partial epilepsy

II.2.A. Objectives
To evaluate short-term efficacy and tolerability during adjunctive therapy use

II.2.B. Primary endpoints
a. Percentage change in seizure frequency during the treatment phase (double-blind phase, including or excluding the dose titration period) compared to baseline.
b. Responder rate (percentage of patients with a greater than 50% reduction in seizures compared to baseline).

II.2.C. Secondary endpoints
a. Percentage change in seizure frequency per seizure type (simple partial, complex partial or secondarily generalized).
b. Percentage of patients with seizure worsening (increase in seizures by 25% or more).
c. Percentage of seizure-free patients.
d. Distribution of responders (seizure reduction of 25 to 50%, 50 to 75% and >75%).
e. Completer rate (measuring the combination of failed efficacy and tolerability).
f. In trials using different dosages, change in seizure frequency and responder rate per dosage group are to be analyzed. Usually, each dose is compared to placebo, but doses are not compared to each other.
g. Incidence and prevalence of adverse events.

**II.2.D. Exploratory endpoints**

a. Relationship between plasma drug concentration and dose, efficacy, and adverse effects. Drug withdrawal and rebound effects.

**II.2.E. Study design**

A multicenter, randomized, placebo-controlled, parallel-group design is generally used. Inclusion criteria usually require stable background treatment with one to three AEDs, a seizure frequency of at least three to four partial seizures per month, and no 28 day periods seizure free. A retrospective screening phase is used to determine the patient’s refractory status, followed by a prospective baseline during which baseline AEDs are held constant and the patient’s seizure diaries are kept. This phase should be of sufficient length to detect fluctuations in seizure frequency (usually 8 to 12 weeks). Shorter baseline periods are preferred, as longer ones retard patient recruitment. Patients are then randomized to placebo and to active treatment (with one or up to three different doses of the study medication). A titration phase of variable length is usually included depending on the characteristics of the drug (this may be crucial because many AEDs are poorly tolerated when started at full maintenance dosages). Titration flexibility might be allowed, i.e. dose may be individually down titrated in cases of poor tolerability with up to two pre-defined lower doses. A maintenance period of 8-16 weeks is then implemented during which treatment remains stable. Concomitant medications are kept stable throughout the trial. If the added agent is expected to modify plasma concentrations of concomitant AEDs, the dosages of the latter may need to be adjusted to maintain constant plasma concentrations. At the end of the trial, the patient is either withdrawn according to a pre-defined withdrawal schedule, or is changed to a predetermined dose of the study medication before unblinding in order to enter an open-label extension trial.

**II.2.F. Planned sample**

Assuming a two-group comparison, a sample size of about 50 to 90 patients per treatment group (depending on the population standard deviation) should permit detection of a difference of 25% between treatment groups in improvement from baseline with 80% power and a type 1 error (two-sided) of 5%.

**II.2.G. Study population**

Adults with refractory partial seizures, with or without secondary generalization, receiving one to three background AEDs

**II.2.H. Specific inclusion criteria**

a. Adults (ages 16 to 65) with partial seizures (simple, complex and secondary generalized tonic-clonic), defined according to the International League Against Epilepsy (ILAE) classification.
b. Patients should have non-controlled seizures despite a stable regimen with 1-3 established appropriate AEDs (the vagal nerve stimulator is sometimes considered as a drug).
c. A defined minimum number of seizures in the baseline phase (e.g., more than 6 observable partial seizures in 8 weeks and at least one seizure in any 4-week period during the baseline).
d. Women of childbearing potential must be using a medically acceptable method of birth control and have a negative serum HCG pregnancy test result at the initial screening visit. Oral contraceptives alone may not be considered adequate because of the potential effect of AEDs on their metabolism.

**II.2.I. Specific exclusion criteria**

a. Patients with generalized epilepsy syndromes.
b. Patients with a history of convulsive status epilepticus in the past year.
c. Patients with non-epileptic attacks (synapses, pseudoseizures).
d. Patients with a clinically relevant medical illness or a significant psychiatric disorder.
e. Patients with progressive CNS disorders (vascular malformations, high grade tumors, etc.).
f. Drug or alcohol abuse.
g. Previous poor compliance with therapy.
h. Pregnant or breastfeeding women.
i. Need for rescue benzodiazepines more frequently than once in the baseline period.
j. Uncountable seizures as a result of seizure clustering, or inadequate supervision if the patient cannot count their own seizures.

II.2.J. Tools for assessing primary endpoints
Seizure diary

II.2.K. Specific criteria for early withdrawal and discontinuation
The investigator will discontinue a patient before completion of the study if consent is withdrawn, if the double-blind code is broken, if it is medically unacceptable to continue treatment due to adverse events, seizure exacerbation, or other reasons, or if pregnancy occurs.

II.2.L. Data analysis method
The analysis of efficacy variables is based on the intention-to-treat (ITT) population (including all randomized patients who have received at least one dose of medication). All statistical tests are two-sided and p values ≤0.05 are considered statistically significant. Multiple statistical methods can be used for the primary and secondary endpoints (analysis of variance, analysis of covariance, logistic regression, Cochran-Mantel-Haenszel statistics, etc).

II.3. Long-term studies
It is customary for patients completing short-term studies to be allowed to enter long-term open-label follow-up studies. The primary objective of these studies is to provide data on tolerability and safety during long-term use and to obtain descriptive information concerning potential loss of therapeutic benefit. Typically, patients entering long-term extension trials may have the dosage of study drug and concomitant medications adjusted on the basis of clinical response. Evaluation will include an overall assessment of tolerability and seizure frequency. The study drug may be continued for as long as it is felt to be clinically beneficial, and retention of patients on the drug may be used as a crude measure of effectiveness. Additionally, extension trials may offer the opportunity to discontinue concomitant medication in order to obtain a preliminary assessment of the study drug under monotherapy conditions.

III. PHASE III STUDIES FOR REGISTRATION OF NEW ANTIEPILEPTIC DRUGS: ADJUNCTIVE THERAPY INDICATIONS

III.1. Outline of a typical development plan
During phase III development at least one large multicenter confirmatory adjunctive therapy, double-blind, randomized, placebo-controlled, parallel group study in refractory partial seizures is performed, usually with a dose-ranging design assessing two or three different doses of the study medication. Adolescents (over 12 years of age) as well as adults may be included in this study. Additional trials in pediatric populations may be performed. These may be conducted in children with refractory partial seizures and, if appropriate based on the expected spectrum of activity of the drug, also in children with refractory generalized epilepsies, such as Lennox-Gastaut syndrome or other syndromes. Monotherapy
trials in refractory patients (conversion to monotherapy design, or presurgical monotherapy design) may also be part of a phase III program: these trials are described in a separate section below.

A clinical development plan will include several different phase III studies, some of which may address special issues (e.g. effects on cognitive function and cognitive outcome, pharmacokinetics and drug interactions, efficacy and tolerability in special groups such as children, the elderly or cognitively impaired patients). The pivotal trial for registration purposes will be an adjunctive therapy trial in refractory partial seizures. Both the FDA and the EMEA will grant approval for the adjunctive therapy indication for partial epilepsy after two adequate and well controlled trials in patients with partial seizures. A trial in Lennox Gastaut syndrome is outlined below to provide an example of a study design for a different epilepsy syndrome.

III.2. Short-term adjunctive therapy studies

Adjunctive therapy trial in refractory partial epilepsy
The design for phase III adjunctive-therapy trials in refractory partial seizures is similar to that described above for phase II trials. Often, multiple doses are explored. Different regimens may also be explored, such as BID versus TID.

Adjunctive therapy trial in Lennox-Gastaut syndrome

III.2. A. Objectives:
To evaluate efficacy and short-term tolerability of adjunctive therapy in patients with Lennox-Gastaut syndrome

III.2.B. Primary endpoints
a. Determination of a statistically significant between-group difference with respect to either 1) reduction in the average monthly seizure rate for all seizure types combined or 2) each component of a compound variable consisting of percentage reduction in drop attacks (tonic-atactic seizures) and the parental global evaluation of seizure severity.
b. Reduction in the average monthly (28-day) seizure rate for all seizure types combined during the treatment phase (double-blind phase, including or excluding the dose titration period) compared to baseline.
c. Percentage reduction in drop attacks (tonic and atonic seizures).

III.2.C. Secondary endpoints
a. Percentages of patients considered to be treatment responders (defined as those with an equal or greater than 50% reduction from baseline for drop attacks, major seizures, and all seizures).
b. Reduction in the average monthly (28-day) rate of major seizures (drop attacks and tonic-clonic seizures).
c. Parental global evaluation of improvement relative to baseline.
d. Incidence and prevalence of adverse events.

III.2.D. Exploratory endpoints
a. Relationship of efficacy and tolerability parameters with dose and plasma drug concentrations. Quantitative EEG analysis.

III.2.E. Study design
A multicenter, randomized, adjunctive therapy, double-blind, placebo-controlled parallel-group design is used. The trial may consist of a 4-week baseline phase followed by a 12-week double-blind treatment phase (including a titration and a maintenance period). Only one dose (weight adjusted) may be explored.
Other aspects of the design are comparable to those described for phase II adjunctive therapy trials in refractory partial epilepsy.

III.2.F. Planned sample
As described above for phase II adjunctive therapy trials in refractory partial epilepsy. Sample size may be moderately lower than in refractory partial epilepsy due to the fact that patients with Lennox Gastaut syndrome tend to have greater seizure frequencies, with a lesser population standard deviation.

III.2.G. Study population
Male and female children (older than 4 years), adolescents and adults with Lennox-Gastaut syndrome

III.2.H. Specific inclusion criteria
a. Adults and children, aged 4 to 65.
   b. History, EEG, and seizure patterns consistent with a diagnosis of Lennox-Gastaut syndrome. Seizure types will include drop attacks (i.e., tonic and atonic seizures) and either a history of or active atypical absence seizures. Other seizure types may include tonic-clonic, myoclonic, and partial-onset seizures. Seizures are classified according to the ILAE International Classification of Epilepsies and Epileptic Seizures.
   c. Patients are required to have refractory and frequent seizures during the month before entering the baseline phase while being maintained on a stable regimen with one or two standard AEDs.

III.2.I. Specific exclusion criteria
a. As described for phase II adjunctive therapy trials in refractory partial epilepsy except, of course, for the epilepsy type.

III.2.J. Tools for assessing primary endpoints
Patient/caretaker seizure diaries, parental global evaluation scale.

III.2.K. Specific criteria for early withdrawal and discontinuation
As described for phase II adjunctive therapy trials in refractory partial epilepsy.

III.2.L. Data analysis method
As described for phase II adjunctive therapy trials in refractory partial epilepsy.

III.3. Long-term adjunctive therapy studies
Please refer to the outline described for phase II long-term adjunctive-therapy studies.

IV. PHASE III STUDIES FOR REGISTRATION OF NEW ANTIPELLEPTIC DRUGS: MONOTHERAPY INDICATIONS

IV.1. Outline of a typical development plan
As monotherapy is the standard treatment for most patients with epilepsy, approval of a monotherapy indication is very important for the success of the drug in the market place. Monotherapy studies also allow evaluation of a drug’s efficacy and tolerability profile, due to removal of the confounding effects of concomitant medication and associated drug-drug interactions.

Because the use of placebo as sole therapy is generally considered ethically unacceptable in epilepsy, most studies use an active control as comparator. This, however, may complicate the interpretation of the results. In fact, when administration of the investigational drug leads to a degree of seizure control which is comparable to that observed with the optimal standard treatment used as a reference (a realistic
scenario, given the remarkable effectiveness of established treatments in the newly diagnosed epilepsy population), the study may be regarded as lacking assay sensitivity, i.e. the two treatments might be equally ineffective in the specific patients’ population recruited in the study. To address these concerns, a number of study designs have been developed which are aimed at demonstrating a difference in favor of the investigational agent. Such protocols involve randomization of patients to a high dosage of the investigational agent and to a suboptimal dosage of either the same agent or an established AED. The use of a suboptimal dose (sometimes referred to as “pseudoplacebo”), however, is controversial as it conflicts with the principle of equipoise which, according to the Declaration of Helsinki, should govern all clinical trials. An additional problem with trials comparing a high versus a low dosage of the investigational agent is that the design is likely to fail to identify the optimal dose range, leading to labeling specifications which may not reflect the optimal mode of use of the drug. For conversion to monotherapy trials, these problems are compounded by the fact that dosage requirements in refractory patients may not necessarily be applicable to patients with newly diagnosed epilepsy, many of whom have milder forms of the disease. With short-term trials, an additional criticism is that the endpoints used and the duration of assessment (see below) may bear little or no relevance to the therapeutic setting, where long-term seizure remission is the major objective to ensure an acceptable quality of life. For the reasons summarized above, regulatory trials which tend to rely on randomization to fixed dosages and relatively short duration of treatment do not provide the information which is required for rational prescribing. Longer duration, flexible-dosages pragmatic trials (see section 5) are suited to address these concerns.

Two different patient populations may be included in monotherapy trials: patients with refractory seizures (usually partial seizures) and patients with newly diagnosed epilepsy. The trial designs used would be different. In refractory patients two types of design are employed: the outpatient conversion to monotherapy and the in-patient presurgical withdrawal to monotherapy. Both designs involve short-term assessment aimed at demonstrating superiority over a suboptimal comparator or placebo. In newly diagnosed epilepsy two types of trials have also been applied: the superiority design, which is usually a medium-term comparison versus a suboptimal comparator or placebo, and the non-inferiority design, which typically involves a longer duration of assessment.

EMEA guidelines for granting the monotherapy indication differ somewhat from those of the FDA. The EMEA requires that the investigational drug should have proven efficacy and safety in newly diagnosed epilepsy, with use in other monotherapy situations being regarded as supportive. Non-inferiority monotherapy trials using an established comparator at optimized dosages are considered by the EMEA as the best study design, even though supportive evidence from some kind of superiority trial (conversion to monotherapy or low-dose vs. high-dose active control) is also recommended. The FDA, on the other hand, does not accept the validity of non-inferiority trials and requires clear demonstration of superiority versus a comparator, either in refractory patients (conversion to monotherapy design) or in newly diagnosed patients. An alternative monotherapy design has been proposed. It involves using historical control data derived from the many withdrawal to monotherapy outpatient trials conducted to date. According to this approach, the investigational agent is assessed at a full dosage in a conversion to monotherapy trial (without including a suboptimal treatment arm, though some other type of control, such as an established AED at a fully effective dosage, may be included for comparative purposes) and a monotherapy license will be granted if the response rate exceeds the upper limit of the confidence interval established for historical controls. In this way, use of a suboptimal treatment and related ethical concerns would be avoided. The validity of this approach is currently being considered by regulatory agencies.

Given this background, a monotherapy development plan aimed at obtaining a worldwide license currently requires at least two separate studies: a superiority trial, conducted preferably in the U.S., and a non-inferiority trial, conducted preferably in Europe.
IV.2. Short-term monotherapy studies

IV.2.i. Outpatient conversion to monotherapy trial in refractory partial epilepsy

IV.2.i. A. Objectives
To evaluate the efficacy and safety of the investigational drug as monotherapy in patients with uncontrolled partial seizures.

IV.2.i. B. Primary endpoints
Time to exit due to fulfillment of one of the exit criteria (the aim is to show that patients allocated to the high dosage of the investigational agent are less likely to experience seizure worsening compared with those allocated to a low-dose suboptimal treatment).

IV.2.i.C. Secondary endpoints
Percentage of patients meeting one of the exit criteria in each of the two treatment groups, incidence and prevalence of adverse events.

IV.2.i.D. Exploratory endpoints
Relationship of efficacy and tolerability parameters with dose and plasma drug concentrations.

IV.2.i.E. Study design
A randomized, double-blind, active control, parallel-group design is used. The trial involves a screening phase and an 8-week baseline phase, which is followed by a treatment phase including a transition period and a monotherapy period. The transition period allows for the treatments being compared to be titrated upwards and for the baseline AEDs to be progressively reduced and eventually discontinued. Withdrawal of background AEDs can be done either before or after randomization. In the more common design, patients are randomized to treatment or control, after which background AEDs are slowly withdrawn over 8-12 weeks. In the second design, all patients are converted to monotherapy treatment with the study drug in an open-label fashion, after which they are randomized to blinded treatment with high vs low dose. The transition phase is followed by a 12- to 16-week monotherapy period that includes an enriched population, i.e. all patients who have successfully converted to monotherapy and did not fulfill exit criteria (see below) in the previous phases. Dose flexibility has sometimes been allowed during this phase. At the end of the treatment phase or if there is a premature discontinuation, the investigational drug is either tapered down and substituted by another AED, or the patient enters an open-label extension phase.

IV.2.i.F. Planned sample
A sample size of about 50 patients per treatment group is required to detect a 35% difference between trial arms in the percentage of patients meeting one exit criteria with 90% power and a type 1 error (two-sided) of 5%.

IV.2.i.G. Study population
Adolescents and adults with refractory partial seizures.

IV.2.i.H. Specific inclusion criteria
Similar to those described for phase II adjunctive therapy trials.

IV.2.i.I. Specific exclusion criteria
Similar to those described for phase II adjunctive therapy trials.

IV.2.i.J. Tools for assessing primary endpoints
Patient’s seizure diary.
**IV.2.i.K. Specific criteria for early withdrawal and discontinuation**
The following exit criteria are defined relative to the number of seizures during the baseline: doubling of average monthly seizure rate, doubling of the highest consecutive 2-day seizure rate, emergence of more severe or new seizure types (including generalized tonic-clonic convulsions), clinically significant prolongation of generalized tonic clonic seizures.

**IV.2.i.L. Data analysis method**
The analysis of efficacy variables is based on the intention-to-treat (ITT) population. Time to exit is analyzed using the log-rank test and Kaplan-Meier estimates.

**IV.2.ii. In-patient presurgical conversion to monotherapy trial in refractory partial epilepsy**

The presurgical design is currently less favored than the conversion to monotherapy design, partly because it has been argued that this design may primarily test efficacy against drug withdrawal seizures, which may involve pathophysiological mechanisms different from those of spontaneous seizures. Moreover, the time scale over which efficacy is assessed in presurgical designs trials is perceived as bearing limited relevance for long-term clinical use. At present, performance of this trial alone would not lead to an FDA indication for use of a drug as monotherapy. It is now more often performed as a proof of principle study, during phase IIa.

This trial requires rapid introduction of the investigational drug, and therefore is only appropriate for compounds that can be initiated rapidly.

**IV.2.ii.A. Objectives**
To evaluate the short-term efficacy and safety of the investigational drug as monotherapy for patients with uncontrolled partial seizures. To prove that a drug enters the brain and has an antiepileptic effect (as proof of principle).

**IV.2.ii.B. Primary endpoints**
Time to exit due to fulfillment of one of the exit criteria. The aim is to show that patients allocated to the high dosage of the investigational agent are less likely to experience worsening of seizures compared with those allocated to a low-dose (suboptimal) treatment.

**IV.2.ii.C. Secondary endpoints**
Percentage of patients completing the study, percentage of patients who meet one of the exit criteria, total number of partial seizures during the double-blind phase, total number of secondarily generalized seizures, incidence and prevalence of adverse events.

**IV.2.ii.D. Exploratory endpoints**
Safety of quick titration/drug-loading, speed of drug action.

**IV.2.ii.E. Study design**
A randomized, double-blind, parallel-group design is used, with a low-dose active control or, at times, a placebo control. The study is performed in patients with refractory partial epilepsy that are admitted to hospital for video-EEG monitoring for presurgical assessment and have their AEDs withdrawn to facilitate seizure recording. Once patients have been drug-free for 48 hours, they are randomized to a high dosage of the investigational drug and to a suboptimal (low-dose or placebo) treatment. To reduce risks, a benzodiazepine (lorazepam) may be permitted during the 48 hour medication-free period and sometimes during the first 24 hours after randomization. The medication or placebo are quickly loaded and the
double-blind evaluation lasts for 8-10 days. Dose flexibility in case of adverse events may be allowed. The double-blind phase may be followed by an open-label extension study.

IV.2.ii.F. Planned sample
Sample size may be calculated with respect to the ability of detecting a 30% difference between the high-dosage group and the placebo/low-dose group for the percentage of patients meeting one exit criteria. If it is assumed that 85% of the placebo-treated patients meet one of the exit criteria, given a two-sided Z-test with a significance level of 0.05 and a power of 0.85, about 50 patients per group are required.

IV.2.ii.G. Study population
Patients with refractory partial seizures.

IV.2.ii.H. Specific inclusion criteria
Patients with refractory partial epilepsy undergoing AED withdrawal within a presurgical workup. Patients need to have between 2-10 partial seizures during baseline and be receiving no AEDs when randomized.

IV.2.ii.I. Specific exclusion criteria
Similar to those described for phase II adjunctive therapy trials in refractory partial epilepsy.

IV.2.ii.J. Tools for assessing primary endpoints
Video-EEG recorded seizures.

IV.2.ii.K. Specific criteria for early withdrawal and discontinuation
Exit criteria are predefined to ensure patient safety. These may include: 3-4 partial seizures or secondarily generalized seizures, new appearance of generalized tonic-clonic seizures, serial/prolonged seizures, or status epilepticus.

IV.2.ii.L. Data analysis method
The analysis of efficacy variables is based on the intention-to-treat (ITT) population. Time to exit may be analyzed using the log-rank test and Kaplan-Meier survival curves. Additional statistical analyses may be performed using a Cox’s proportional hazards regression model. Secondary efficacy variables (percentage of patients who meet one of the exit criteria) may be analyzed using the Cochran-Mantel-Haenszel test.

IV.3. Long-term monotherapy studies

IV.3.i. Superiority monotherapy trial in newly diagnosed epilepsy

IV.3.i.A. Objectives
To evaluate the comparative efficacy and tolerability of the investigational drug versus a (usually suboptimal) active control under monotherapy conditions in new onset epilepsy.

IV.3.i.B. Primary endpoint
Time to first or second seizure (seizures occurring during the titration period may or may not be censored, depending on the characteristics of the titration phase).

IV.3.i.C. Secondary endpoints
Time to treatment failure (discontinuation of treatment), percentage of seizure-free patients after 6 and 12 months of treatment.
IV.3.i.D. Exploratory endpoints
Relationship of efficacy and tolerability parameters with dose and plasma drug concentrations.

IV.3.i.E. Study design
A multicenter, randomized, double-blind, parallel-group study is performed in patients with untreated epilepsy using a dose-controlled design, i.e. comparing a low dosage with a high dosage. Target dosages may be reached after an appropriate titration period, and patients experiencing intolerable adverse effects during titration may be allowed to step back by one dose level. The aim of the study is to demonstrate that time to first or second seizure is longer in the high-dosage group than in the low-dosage group. The double-blind phase may be followed by an open-label extension study.

IV.3.i.F. Planned sample
It has been suggested that these studies be powered on the basis of number of failure events (a first seizure), not number of patients. Based on the hypothesis that the hazard ratio for time to first seizure is 0.525 and constant over time, 108 events are needed for 92.5% power to detect a statistically significant difference at the 5% (two-sided) level. Enrollment is to be stopped when 108 events have been observed. In the only trial performed using this approach and only allowing the inclusion of patients with 1 or 2 seizures during a retrospective 3-month baseline, a recruitment of about 500 patients was necessary.

IV.3.i.G. Study population
Patients with new onset epilepsy or previously diagnosed but currently untreated epilepsy. The population could include patients with partial seizures (with or without secondary generalization) and primarily generalized tonic-clonic seizures.

IV.3.i.H. Specific inclusion criteria
a. Age of 12 to 65 years (wider age limits may be acceptable).
b. Recently diagnosed epilepsy with two or more unprovoked seizures. It may be possible to accept one seizure plus additional unequivocal evidence supporting the diagnosis of epilepsy (epileptiform EEG activity, brain imaging evidence). Patients should have had at least one seizure within the 3 months previous to randomization (an upper limit to number of seizures during this period may be set). In addition, patients with a previous history of epilepsy that has been in remission without medications for at least 6 months and have had one seizure in the previous 3 months may be included.
c. Patients treated with a single AED for less than 2 weeks could enter the study provided that medication is withdrawn previous to randomization.
d. Women of childbearing potential must be using a medically acceptable method of birth control and have a negative serum HCG pregnancy test result at initial screening visit.

IV.3.i.I. Specific exclusion criteria
Non-epileptic attacks (syncope, pseudoseizures), history of status epilepticus, significant medical or psychiatric illness, drug abuse, progressive central nervous system disease. Depending on the drugs being compared, patients with specific epilepsy syndromes (for example, generalized epilepsies) may need to be excluded.

IV.3.i.J. Tools for assessing primary endpoints
Seizure diary.

IV.3.i.K. Specific criteria for early withdrawal and discontinuation
First seizure or adverse event requiring discontinuation of treatment.
IV.3.i.i.L. Data analysis method
The analysis of efficacy variables is based on the intention-to-treat (ITT) population. The primary efficacy variable is analyzed by Kaplan-Meier survival analysis. The log-rank test may be used to assess between-group differences. A Cox proportional hazards model may also be applied.

IV.3.ii. Sequential design monotherapy trial in newly diagnosed epilepsy

IV.3.ii.A. Objectives
To evaluate the comparative efficacy and tolerability of the investigational drug versus an active control (an established AED or a lower dose of the investigational drug) under monotherapy conditions in new onset epilepsy.

IV.3.ii.B. Primary endpoints
Time to first seizure following completion of the dose titration phase.

IV.3.ii.C. Secondary endpoints
Time to treatment failure (discontinuation of treatment), percentage of seizure-free patients after 6 and 12 months of treatment. Time to second, third and fourth seizure.

IV.3.ii.D. Exploratory endpoints
Relationship of efficacy and tolerability parameters with dose and plasma drug concentrations.

IV.3.ii.E. Study design
The sequential design is a trial which allows a series of interim analyses of the emerging data so that the trial can be stopped when a predetermined difference between treatments (or lack of such a difference) has been demonstrated. These trials may require fewer patients than traditional designs of equal power, and in particular can avoid continuation when one treatment is already evidently inferior to the other. As with other trials, the design involves a multicenter, randomized, double-blind, active control, parallel-group comparison. In the only large trial where this design was used in epilepsy, patients were titrated up to a predetermined target dosage and thereafter followed up for a maximum of 162 weeks. Dose adjustments were permitted if seizure control was inadequate or adverse events were observed. The double-blind phase may be followed by an open-label extension study.

IV.3.ii.F. Planned sample
The main trial conducted to date utilized a prediction of survival curves for the reference treatment (carbamazepine, initial target dose 600 mg/day) and for the investigational drug. It was regarded as possible that the investigational drug would improve the probability of surviving without a study event for 54 weeks from 0.5 to 0.6, a difference considered clinically significant. Based on the model used, a total sample size of 450 to 700 recruited patients (depending on the true seizure rates on the two treatments) was expected to be necessary to demonstrate the target difference with a power of 0.90 and at the 5% level (two-sided). This would compare with a sample size of about 1,000 patients if a fixed (non-sequential) sample design of equal power is used.

IV.3.ii.G. Study population
Patients with new onset epilepsy or previously diagnosed but currently untreated epilepsy. The population could include patients with partial seizures (with or without secondary generalization) and with primarily generalized tonic-clonic seizures (if experimental drug is broad-spectrum).

IV.3.ii.H. Specific inclusion criteria
As described above for non-sequential design monotherapy trials in newly diagnosed epilepsy.
IV.3.ii.I. Specific exclusion criteria
As described above for non-sequential design monotherapy trials in newly diagnosed epilepsy.

IV.3.ii.J. Tools for assessing primary endpoints
Seizure diary.

IV.3.ii.K. Specific criteria for early withdrawal and discontinuation
Uncontrolled seizures at the highest dosage allowed by the protocol, or adverse event requiring discontinuation.

IV.3.ii.L. Data analysis method
Each interim analysis may comprise a comparison of the survival rates on the two treatments by means of Cox’s proportional hazards regression, adjusting for seizure type and for the number of seizures during the 12 months prior to randomization. The statistics assessing the advantage of one of the treatments is denoted by the Z score, which generalizes the better known log-rank statistics to allow for any imbalance in prognostic factors. Additionally, a measure of information, denoted by V, is calculated as the null variance (approximately equal to one quarter of the total number of events). These statistics are plotted against each other at each data review, until one of the stopping boundaries of the design is crossed.

IV.3.iii. Non-inferiority monotherapy trials

IV.3.iii.A. Objectives
To evaluate the medium to long term efficacy and tolerability of an investigational drug in patients with newly onset epilepsy in comparison with an established licensed in monotherapy AED at fully effective dosages.

IV.3.iii.B. Primary endpoints
   a. Proportion of patients seizure-free for 6 months assessed in the per-protocol (PP) population.

IV.3.iii.C. Secondary endpoints
   a. Proportion of patients seizure-free for 6 months assessed in the intention-to-treat (ITT) population.
   b. Proportion of patients seizure-free for 6 months in a subset of the per-protocol (PP) population which excludes drop-out for reasons unrelated to efficacy.
   c. Percentage of patients who remain seizure-free for 12 months.
   d. Time to exit.
   e. Percentage of completers.
   f. Time to first or second seizures.
   g. Percentage of patients seizure-free at each dose.
   h. Percentage of patients withdrawn due to adverse events.

IV.3.iii.D. Exploratory endpoints
   a. Relationship of efficacy and tolerability parameters with dose and plasma drug concentrations, cognitive function measures, quality of life measures.

IV.3.iii.E. Study design
The trial may involve a multicenter, double-blind, randomized, parallel-group design comparing the investigational drug with the best reference treatment at optimized dosages. Patients are allocated to an initial target dosage of both drugs at the lower end of the expected optimal range. If the primary endpoint (6-month seizure-freedom) is not reached due to seizure recurrence after the target dose has been attained, the patient is up-titrated to a higher pre-determined dosage. If the primary endpoint (6-month seizure-freedom) is again not reached due to seizure recurrence, the patient is up-titrated to the highest dosage
level. Patients may be allowed to step back to an intermediate dosage if side effects are encountered during each of the titration phases. The double-blind phase may be followed by an open-label extension study.

**IV.3.iii.F. Planned sample**
Assuming a 6-month seizure-free rate of 45% with the reference comparator, a true difference of zero between treatments and a 20% rate of protocol violators, a total sample size of 580 recruited patients would be required to ensure a lower limit above –15% for the two sided 95% confidence interval for the difference in 6-month seizure-free rates, with 90% power.

**IV.3.iii.G. Study population**
Patients with new onset epilepsy or previously diagnosed but currently untreated epilepsy. The typical population could include patients with partial seizures (with or without secondary generalization) and with primarily generalized tonic-clonic seizures, if the investigational drug is broad-spectrum.

**IV.3.iii.H. Specific inclusion criteria**
As described above for non-sequential design superiority monotherapy trials in newly diagnosed epilepsy.

**IV.3.iii.I. Specific exclusion criteria**
As described above for non-sequential design superiority monotherapy trials in newly diagnosed epilepsy.

**IV.3.iii.J. Tools for assessing primary endpoints**
Seizure diary

**IV.3.iii.K. Specific criteria for early withdrawal and discontinuation**
Seizures uncontrolled at the highest dosage level, adverse events requiring discontinuation of treatment.

**IV.3.iii.L. Data analysis method**
In non-inferiority trials, analysis of the primary efficacy variable is made on the per-protocol (PP) population. Six-month seizure-free rates may be compared by a logistic regression model whose 95% confidence interval computation may include treatment and seizure types (e.g. partial vs. generalized tonic-clonic seizures without clear focal onset) as factors. Interactions between treatment group and seizure types may be excluded by applying an additional logistic regression model including treatment, seizure type and treatment by seizure type factors.

**V. OTHER STUDIES (SPECIAL INDICATIONS, PRAGMATIC TRIALS)**

**V.1. Paediatric epilepsies**
Studies of AEDs in children are conducted in three main patient populations:

a. Patients with refractory partial seizures. For this population, a multicenter, randomized, double-blind, adjunctive therapy, parallel-group, placebo-controlled trial may be performed using a design similar to that described for adults;

b. Patients with refractory generalized epilepsy. For this population, a multicenter randomized, double-blind, adjunctive therapy, parallel-group, placebo-controlled trial similar to that described in section 3 for Lennox-Gastaut syndrome may be performed. Studies in certain syndromes may require specific protocols: for example, trials in absence epilepsy may be of shorter duration and should use EEG changes (e.g., reduction/disappearance of spike-and wave activity) as primary efficacy endpoint;

c. Patients with newly diagnosed partial or generalized epilepsy. These studies usually involve multicenter, randomized, double-blind, monotherapy, active control trials, which are initiated after evidence of efficacy has been obtained from adjunctive therapy trials. The range of designs is
similar to those described for monotherapy trials in adults. Not uncommonly, inclusion criteria for monotherapy trials in newly diagnosed patients allow simultaneous inclusion of children and adults. Studies in certain paediatric epilepsy syndromes, however, may require syndrome-specific protocols.

Both the FDA and the EMEA indicate that safety data in paediatric populations should be included in the registration dossier. The EMEA suggests that paediatric studies should be initiated as early as the development program allows and that a minimum of 100 children should be followed-up for at least one year. Moreover, it is recommended that short-term and long-term studies be designed to detect possible impact on learning, intelligence, growth, endocrine functions and puberty. Paediatric pharmacokinetic data are also required.

V.2. Epilepsies in the elderly

Pharmacokinetic and safety data in a reasonable number of elderly patients (100 or more) should be collected during phase III. This should include an evaluation of potential effects on cognitive function and sedation, as well as interactions with medications frequently used in this age group. Randomized monotherapy trials in elderly patients may be performed to obtain information supporting the use of a new AED in this segment of the population.

V.3. Acute repetitive seizures and status epilepticus

Acute repetitive seizures and status epilepticus are emergency situations where acute treatment, usually by the parenteral route, is indicated. Special trial designs are required for these indications. These involve multicenter randomized active control trials, though in some circumstances (which exclude convulsive status epilepticus) use of placebo may be justifiable. Examples of trial designs for these conditions can be found among the landmark trials listed in section 7.

V.4. Pragmatic trials

Pragmatic trials are designed to reproduce conditions which more closely reflect the use of a drug in routine clinical practice. Randomized pragmatic trials may be designed to assess the relative value of different therapeutic strategies rather than individual drugs (for example, early versus deferred treatment in those situations where the indication to treat is in doubt) or to compare two or more AEDs under conditions which allow physicians to optimize dosages and other treatment modalities according to personal clinical judgement. These studies may be of little value to regulators, but they provide useful information on which to guide rational prescribing. Most of these trials follow a design similar to the randomized non-inferiority active-control monotherapy trial outlined in section 4, but they allow greater dosing flexibility and use more relevant clinical endpoints (e.g., 12-month seizure remission rates, and retention on the allocated treatment). One example of such studies is the landmark Veterans Administration trial which compared phenobarbital, primidone, carbamazepine and phenytoin in patients with partial and/or secondarily generalized tonic-clonic seizures.

VI. EXAMPLES OF LANDMARK WELL DESIGNED TRIALS

Adjunctive therapy, partial epilepsy

**Adjunctive therapy, Lennox-Gastaut-syndrome**

**Monotherapy, short-term trials**

**Monotherapy, long-term trials**

**Acute repetitive seizures and status epilepticus**

**VII. SUGGESTED READINGS**
   c. E10 choice of control group and related issues on clinical trials. 2001 (http://www.fda.gov/cder/guidance/4155fml.htm)
Chapter 17. Headache Disorders

T.J. Steiner, Ph.D. ¹

Peer Tfelt-Hansen, M.D., Ph.D. ²

Douglas C. McCrory, M.D. ³

¹ Division of Neuroscience
Imperial College London
St. Dunstan’s Road
London W6 8RP
UK

² Department of Neurology
University of Copenhagen
Glostrup Hospital
DK-2600 Glostrup
DENMARK

³ Center for Clinical Health Policy Research
Department of Medicine
Duke University Medical Center and the Center for Health Services Research in Primary Care
Durham Veterans Affairs Medical Center
Durham, NC
USA
I. INTRODUCTORY REMARKS

Headache disorders are ubiquitous. Their lifetime prevalence in populations in which it has been measured is over 90%. They are also disabling, and costly.

*Migraine* manifests in attacks lasting a few hours to three days, with a median frequency of one per month. It has a 1-year prevalence of 12-15% throughout Europe and in North and South America. In other parts of the world there is gathering evidence of similar prevalence. In Japan it is estimated to affect 8.4% of the adult population. Elsewhere in the Far East, surveys of representative samples of the general population are difficult; prevalence and impact, which have been thought to be lower, may be underestimated. In an American survey 80% of people with migraine reported some degree of disability because of it and globally, according to WHO’s *World Health Report 2001*, migraine is 19th in the list of all causes of years lived with disability. Because migraine affects people particularly during their productive years, its economic impact is high. Additionally, as well as suffering directly from its symptoms, people with migraine consistently score highly on scales of general physical and mental ill-health.

*Tension-type headache* is the most common headache disorder. Most is episodic, with occasional attacks lasting hours of what sufferers often describe as “ordinary headache”. In its chronic and much more disabling sub-type, on the other hand, it is present on more days than not. *Chronic tension-type headache* overlaps with and is sometimes indistinguishable from other forms of *chronic daily headache*, some of which are unrelentingly present throughout every day. Estimates of the prevalence of this group of conditions in Europe and USA are as high as 1 in 25 of the *entire* adult population. Such frequent headache is associated with long-term morbidity and disability.

*Cluster headache* has a lower prevalence (lifetime about 0.07%) and at any given time most people with the disorder are not in a cluster period. It is unique amongst the primary headache disorders in affecting men more than women (about 6:1). Typically occurring in bouts of a few weeks each year with periods of full remission between, it is characterised by frequent (daily or more often) short-lasting (15-120 minutes) but excruciating unilateral localised frontal or periorbital pain accompanied by marked but similarly localised autonomic symptoms. In its rarer chronic sub-type there are no periods of remission.

The financial cost of headache arises partly from direct treatment costs but, to a much greater extent, from consequential losses: work time and productivity losses are by far the largest elements. That these costs remain high throughout the world is evidence of treatment failure, a problem attributable in part to the fact that available drugs fall well short of being ideal treatments for any of these headache disorders. In another important part, it is due to the low priority generally given to headache in the queue for health-care resource allocation.

Drugs are therefore required that are not only more effective but also shown to be cost-effective in the relief of headache. Surprisingly, pharmacoeconomic studies in this area are in their infancy. Whilst research is needed to derive simple agreed cost-of-illness measures that adequately capture those aspects of cost that matter to patients, time is an important casualty of headache and time losses (and their reduction by effective treatment) should be relatively easy to measure.

Amongst the primary headaches, only migraine has benefited substantially from recent pharmaceutical investment. The result has been the marketing since 1991 of seven triptans, a class of drug which has unquestionably moved forward the acute treatment of migraine whilst proving to be of some value, albeit limited, in cluster headache also. Yet triptans are far from being 100% efficacious, and they are not universally tolerated. Whilst much current research concentrates on head-to-head comparisons between triptans, invariably showing minor differences, development of new drug classes may have stalled. Meanwhile, symptomatic treatments (analgesics and anti-emetics) remain very useful in acute migraine
therapy. They are still the mainstay of treatment of migraine in children and for many adults, including those for whom triptans are not appropriate, and they are the only option in the large parts of the world that new and relatively expensive drugs do not penetrate. They are, generally, the first-line treatments for tension-type headache.

The triptan successes displaced interest from preventative studies. For all primary headache disorders, prophylactic drugs now available are, at best, of quite limited value.

Clearly there is much unfulfilled therapeutic need in all of these areas requiring further clinical research. For pharmaceutical companies, potential markets are very large indeed although experience shows that the greater part of these is not easily penetrated on the basis of efficacy studies alone.

II. PHASE II STUDIES FOR REGISTRATION OF NEW DRUGS

II.1. Outline of a typical development plan

Whilst the requirements differ for migraine, tension-type headache and cluster headache, and in each of these for acute and preventative therapies, there are general principles applying to all.

During phase II, candidate drugs are assessed for efficacy against placebo. Randomised controlled trials will follow pharmacokinetic studies. For drugs intended for the acute treatment of migraine, pharmacokinetic studies must be conducted during the attack because absorption may be delayed by gastric stasis. For this reason oral administration, though preferred by most patients, is not ideal in acute migraine: early proof-of-concept studies may use parenteral therapy if there is doubt about rapid bioavailability by the oral route and alternative routes may anyway be required in severely nauseated or vomiting patients. Cluster headache attacks typically have duration of 15-120 minutes; rapid bioavailability is needed if treatment is to have worthwhile effect, and this is unlikely to be achieved by the oral route. Phase II must address formulation issues.

Recent experience in migraine has shown that phase II often needs to incorporate dose-finding studies, and the EMEA now require to know both the lower end of the clinically effective dose range and the optimal dose(s) (inter-individual variation may be high enough that a range of doses should be marketed). Typical studies will be double-blind, involving one or up to three or more doses. They may use parallel-groups or (multiple) cross-over designs, with regulators generally favouring the former.

Regulators have not so far distinguished, for drug-development purposes, between episodic tension-type headache and other causes of mild-to-moderate pain.

Acute-treatment trials in phase II usually allow the treatment of a single attack. Attrition rates tend to be high between attacks (one of the arguments against cross-over designs).

Prophylactic trials must allow time for dose-titration (if needed) and then time for effect to develop and be measurable, usually requiring a minimum of three months’ treatment for migraine or chronic tension-type headache. The objectives of treatment in these two disorders are not identical: migraine prophylaxis is intended to reduce the frequency of continuing attacks; in chronic tension-type headache the intent is to cause reversion to the episodic subtype. Patients with cluster headache, where use of placebo as a control may be difficult, will expect rapid and obvious efficacy achieving attack suppression or, ideally, attack cessation. Furthermore, in episodic cluster headache, spontaneous remission of the cluster period attenuates a trial’s ability to detect true treatment benefits. Any need for a period of dose-titration to balance efficacy and tolerability is a significantly complicating factor in cluster headache trial design. Long-term continuation protocols in cluster headache are not part of phase II.
Most headache disorders are appropriately treated in primary care and arguably that is where studies should be done, but this may not be true of phase II. Nevertheless, at this stage of development patients should be selected who are typical of the disorder, particularly avoiding those with complicated or refractory presentations who tend to accumulate in specialist headache centres. Exclusion criteria commonly markedly constrain recruitment, and in most cases multiple centres are needed to reach recruitment targets within reasonable timeframes.

Pharmacokinetic interactions between the test agent and other drugs likely to be taken in clinical practice should be evaluated early on. The majority of headache patients are women of child-bearing potential so oral contraceptives are important amongst these. Teratogenicity should also be assessed early: not only will teratogenic drugs be of little value in headache but also, with the notable exception of cluster headache, it is extremely difficult to recruit to headache studies if women of child-bearing potential must be excluded.

II.2. Short term studies

II.2.i. Acute treatment of migraine

II.2.i.A. Objectives

To evaluate efficacy in relieving symptoms of the acute attack.

II.2.i.B. Primary end-point

a. “Pain-free” rate: percentage of patients pain-free at 2 hours after treatment.

II.2.i.C. Secondary endpoints

a. “Headache-relief” rate: percentage of patients with a decrease in headache intensity from severe or moderate to mild or no pain within 2 hours after treatment (this was widely adopted as the primary end-point in past studies and remains important in phase II for purposes of comparison).

b. Percentage of patients pain-free at 2 hours and remaining pain-free, without use of further medication, for 48 hours after treatment.

c. Rate of relapse, defined as the return of headache of any intensity within 48 hours in patients pain-free at 2 hours after treatment.

d. Headache intensity at various time points after treatment.

e. Functional disability on a validated scale (usually a 4-point verbal rating scale where 0 = no functional impairment, 1 = “can do everything albeit with difficulty”, 2 = “cannot do some things” and 3 = “cannot do anything and/or bed-bound”) at 2 hours and other time points after treatment.

f. Effect on associated symptoms such as nausea, vomiting, photophobia and phonophobia.

g. Rate and timing of use of rescue medication.

h. Incidence and nature of adverse events.

II.2.i.D. Study design

These are invariably short-term studies. Recommended are randomised, double-blind, placebo-controlled parallel-groups studies treating one attack per patient. Three-arm trials, including placebo, are required for internal validation in active-comparator studies because of the large and highly variable placebo effect in migraine studies. In general, there is no need for stratification. Treatment may, depending on its nature, require the patient to attend a treatment facility such as the doctor’s office or be taken by the patient at home or wherever he or she may be. The former is logistically difficult. In the latter case, outcome variables are usually recorded by the patient in paper or electronic diaries, with prompts at various time points. Rescue medication should be allowed after 2 hours (for this reason, attack duration is not considered a useful secondary endpoint). The observation period after treatment of an attack should be 48 hours. Outpatients should return for final review soon after this.
II.2.i.E. Planned sample
Sample size calculations should be based on demonstrating superiority over placebo in a primary analysis of difference in pain-free rates, with an absolute difference of 20% being clinically significant and allowing for a placebo rate of up to 20%.

II.2.i.F. Study population
Adults with migraine with or without aura.

II.2.i.G. Specific inclusion criteria
a. Patients with migraine conforming to IHS diagnostic criteria 1.1 or 1.2 for at least 1 year and with at least 3 months’ well-documented retrospective history.
b. Migraine attacks occurring 1-6 times monthly.
c. Males and females.
d. Unless otherwise justified, patients should be over 18 years of age.

At the time of treatment:
a. An acute attack, usually with onset within the previous 12 hours.
b. At least 48 hours since resolution of the previous attack.
c. Headache of moderate or severe intensity.
d. So far untreated.

II.2.i.H. Specific exclusion criteria
a. Age at onset of migraine of 50 years or over.
b. Other headaches not well distinguished from migraine or occurring with such frequency as to interfere with assessments (usually taken to mean occurring on >6 days/month).
c. Other illnesses likely to interfere with assessments.
d. Use of migraine prophylactic drugs in the previous month.
e. Use of or requirement for other unacceptable concomitant therapy.
f. History of drug or alcohol overuse.

II.2.i.I. Tools for assessing endpoints
Paper or electronic diaries, with prompts at various time points.

II.2.i.J. Data analysis method
In early efficacy studies, explanatory (per protocol) analysis may be appropriate. It is unhelpful at this stage to include patients with major protocol violations. Standard statistical methods are appropriate. Adverse events are usually analysed descriptively.

II.2.ii. Acute treatment of episodic tension-type headache

II.2.ii.A. Objectives
To evaluate efficacy in relieving pain and functional impairment attributable to acute episodic tension-type headache.

II.2.ii.B. Primary end-point
a. “Pain-free” rate: percentage of patients pain-free at 2 hours after treatment.

II.2.ii.C. Secondary endpoints
a. Headache intensity (scored on either a visual analogue scale or a 4-point verbal rating scale [0 = no pain; 1, 2, 3 = mild, moderate, severe pain]) at 2 hours and other time points after treatment.
b. Headache intensity difference (the arithmetic change from baseline in headache intensity score) at 2 hours and at other time points after treatment.
c. Headache relief (on a verbal rating scale from “none” to “complete”, with two or more intermediaries which may include “meaningful relief”; negative scores may be incorporated to indicate worsening) at 2 hours and at other time points after treatment.
d. Functional disability on a validated scale (e.g., a 4-point verbal rating scale where 0 = no functional impairment, 1 = “can do everything albeit with difficulty”, 2 = “cannot do some things” and 3 = “cannot do anything and/or bed-bound”) at 2 hours and other time points after treatment.
e. Rate and timing of use of rescue medication.
f. Incidence and nature of adverse events.

II.2.ii.D. Study design
These are invariably short-term studies. Recommended are randomised, double-blind, placebo-controlled parallel-groups studies treating one attack per patient. Three-arm trials, including placebo, are required for internal validation in active-comparator studies because of the very large placebo effect reported in acute episodic tension-type headache studies. There is no need for stratification. Treatment is taken by the patient at home or wherever he or she may be. Outcome variables are usually recorded by the patient in paper or electronic diaries, with prompts at various time points. Rescue medication should be allowed after 2 hours. The observation period after treatment should be at least 24 hours. Patients should return for final review soon after this.

II.2.ii.E. Planned sample
Sample size calculations should be based on demonstrating superiority over placebo in a primary analysis of difference in pain-free rates, with an absolute difference of 20% being clinically significant and allowing for a placebo rate of up to 50%.

II.2.ii.F. Study population
Adults with episodic tension-type headache drawn (by advertising if necessary) from the general population (this is not a disorder that usually causes medical consultation; if it does, this is probably because of complicating factors or comorbidity).

II.2.ii.G. Specific inclusion criteria
   a. Patients with frequent episodic tension-type headache (occurring on >1 but <15 days per month) conforming to IHS diagnostic criteria 2.2 for at least 1 year and with at least 3 months’ well-documented retrospective history.
   b. Usual headache duration at least 4 hours.
   c. Males and females.
   d. Unless otherwise justified, patients should be over 18 years of age.

At the time of treatment:
   a. An acute episode of tension-type headache, usually with onset within the previous 12 hours.
   b. Headache of at least moderate intensity.
   c. So far untreated.

II.2.ii.H. Specific exclusion criteria
   a. Age at onset of tension-type headache of 50 years or over.
   b. Chronic tension-type headache.
   c. Other headaches, especially migraine and medication-overuse headache.
   d. Other illnesses likely to interfere with assessments.
   e. Use of prophylactic drugs in the previous month.
f. Use of or requirement for psychotropic medication or other unacceptable concomitant therapy.
g. History of drug or alcohol overuse.

II.2.ii.I. Tools for assessing endpoints
Paper or electronic diaries, with prompts at various time points.

II.2.ii.J. Data analysis method
In early efficacy studies, explanatory (per protocol) analysis may be appropriate. It is unhelpful at this stage to include patients with major protocol violations. Standard statistical methods are appropriate. Adverse events are usually analysed descriptively.

II.2.iii. Acute treatment of episodic or chronic cluster headache

II.2.iii.A. Objectives
To evaluate efficacy in aborting or suppressing the acute attack.

II.2.iii.B. Primary end-point
a. “Aborted attack” rate: percentage of patients in whom the attack is effectively stopped (headache intensity reduced to mild or no pain) within a prescribed time interval (which may be as short as 10 minutes).

II.2.iii.C. Secondary endpoints
a. Time to meaningful relief.
b. Time to “complete” relief (mild or no pain).
c. Rate of relapse, defined as the return of headache of moderate or greater intensity within 1 hour in patients reporting an aborted attack.
d. Rate and timing of use of rescue medication.
e. Incidence and nature of adverse events.

II.2.iii.D. Study design
The dramatic nature of cluster headache attacks and low placebo-response rates make open and single-blind trials informative as pilot studies. Formal comparisons with placebo must follow. These are invariably short-term randomised, double-blind, placebo-controlled studies with parallel-groups or crossover design. The latter may be preferable in this disorder since patients are uncommon but attacks occur with high frequency. In phase II, episodic and chronic cluster headache should probably be separated; if they are not, stratification is recommended in parallel-groups studies because responses to treatment may differ. Stratification for gender is also recommended for the same reason. Each patient should treat or be treated for one attack with study medication. Treatment may, depending on its nature, require the patient to be admitted to hospital, or attend a treatment facility such as the doctor’s office at a time when an attack is anticipated, or it may be taken by the patient at home or wherever he or she may be. Outcome variables are usually recorded by the doctor or patient in paper or electronic diaries, with prompts at various time points. Rescue medication should be allowed after the time prescribed for the primary end-point, but options for this are very limited. The observation period after treatment of an attack should be not less than 24 hours unless interrupted by the occurrence of the next attack. Outpatients should return for final review within 2 days.

II.2.iii.E. Planned sample
Sample size calculations should be based on demonstrating superiority over placebo in a primary analysis of difference in aborted-attack rates. Placebo-response rate is low but, because of the short attack-duration, spontaneous remission rates may be high; the two may combine to 50%. An absolute difference of 25% is clinically significant.
II.2.iii.F. Study population
Adults with episodic or chronic cluster headache.

II.2.iii.G. Specific inclusion criteria
a. Patients with episodic or chronic cluster headache conforming to IHS diagnostic criteria 3.1 or 3.2; patients with episodic cluster headache should be in at least their second cluster period.
b. Acute attacks occurring between once every 2 days and 5 times per day.
c. Attack duration of 30-180 minutes.
d. Males and females.
e. Unless otherwise justified, patients should be over 18 years of age.

At the time of treatment:

a. An acute attack, usually with onset within the previous 15 minutes (at least 15 minutes before expected spontaneous resolution).
b. At least 1 hour since resolution of the previous attack and 24 hours (or 5 half-lives if longer) since the latest previous use of study drug.
c. Headache of moderate or greater intensity.
d. So far untreated.

II.2.iii.H. Specific exclusion criteria
a. Other headaches not well distinguished from cluster headache.
b. Other illnesses likely to interfere with assessments.
c. Concurrent use of prophylactic drugs for cluster headache.
d. Use of or requirement for other unacceptable concomitant therapy.
e. History of drug or alcohol overuse.

II.2.iii.I. Tools for assessing endpoints
Paper or electronic diaries, with prompts at various time points.

II.2.iii.J. Data analysis method
In early efficacy studies, explanatory (per protocol) analysis may be appropriate. It is unhelpful at this stage to include patients with major protocol violations. Subgroup analyses (for episodic and chronic subtypes and for gender differences) are recommended and should be specified a priori. Standard statistical methods are appropriate. Adverse events are usually analysed descriptively.

II.3. Long term studies

II.3.i. Migraine prophylaxis

II.3.i.A. Objectives
To evaluate efficacy in migraine-attack prevention.

II.3.i.B. Primary end-points
a. Frequency of attacks per specified unit time (usually 4 weeks) measured during treatment after a specified period (usually 8 weeks).
b. Response rate: percentage of patients with frequency reduction of 50% or more after a specified treatment period.

The number of attacks should be recorded irrespective of their duration, and the following rules distinguish an attack of long duration from two attacks and between attacks and relapses:
a. A migraine attack which is interrupted by sleep, or which temporarily relents spontaneously and then recurs within 48 hours after its onset, should be recorded as one attack and not two.
b. An attack treated successfully with medication but with relapse within 48 hours counts as one attack.

II.3.i.C. Secondary endpoints
a. Frequency of attacks over the entire treatment period.
b. “Migraine days” (defined as any day on which symptoms of migraine are present) per 4 weeks.
c. Intensity of migraine headache averaged over attacks within a specified evaluation period.
d. Drug consumption for symptomatic or acute treatment totalled over an evaluation period.
e. Incidence and nature of adverse events.

II.3.i.D. Study design
These are invariably medium-term studies (at least 4 months) conducted in outpatients. Recommended are randomised, double-blind, placebo-controlled parallel-groups studies. Three-arm trials, including placebo, are required for internal validation with active comparators because of the large and highly variable placebo effect in prophylactic migraine studies. Randomisation should occur after a run-in (baseline) period of at least one month, when stratification for baseline attack rate (e.g., ≥3 or <3 per 4 weeks) is recommended as the prophylactic effect may depend on this variable. Treatment periods should be at least 3 months. Patients should take their usual acute therapy as required, provided that it can be safely administered with the study drug. Attacks (and, if required, their features), acute medication use and adverse events should be recorded as they occur by patients in paper or electronic diaries. Reviews of patients and their diaries should be undertaken every 4 weeks.

Compliance with medication regimens, and concordance, are known to be highly suspect in migraine prophylaxis. Counts of returned medication are unreliable for detecting poor concordance, which may render an efficacious drug useless. In phase II it is especially important to ascertain that the drug has been taken as prescribed. Consideration should be given to using electronic event monitors.

II.3.i.E. Planned sample
Sample size calculations should be based on demonstrating superiority over placebo in a primary analysis of a) difference in attack frequencies, with a relative difference of 50% or an absolute difference of 1 attack/month being clinically significant and allowing for a reduction on placebo of up to 30% or 1 attack/month; or b) difference in responder rates, with an absolute difference of 20% being clinically significant and allowing for a placebo rate of up to 30%.

II.3.i.F. Study population
Adults with frequent attacks of migraine with or without aura.

II.3.i.G. Specific inclusion criteria
a. Patients with migraine conforming to IHS diagnostic criteria 1.1 or 1.2 for at least 1 year and with at least 3 months’ well-documented retrospective history.
b. Migraine attacks occurring 2-6 times monthly.
c. Males and females.
d. Unless otherwise justified, patients should be over 18 years of age.

II.3.i.H. Specific exclusion criteria
a. Age at onset of migraine of 50 years or over.
b. Other headaches not well distinguished from migraine or occurring with such frequency as to interfere with assessments (usually taken to mean occurring on >6 days/month).
c. Other illnesses likely to interfere with assessments.
d. Use of other migraine prophylactic drugs in the previous month.
e. Use of or requirement for other unacceptable concomitant therapy.
f. Risk of pregnancy.
g. History of drug or alcohol overuse.

II.3.i.I. Tools for assessing endpoints
Paper or electronic diaries.

II.3.i.J. Specific criteria for early withdrawal and discontinuation
a. Withdrawal of consent (which may be related to lack of efficacy or adverse events).
b. Medical concerns related to evident lack of efficacy or adverse events.
c. Other intercurrent illness.
d. Pregnancy.
e. Breach of double-blinding.

II.3.i.K. Data analysis method
Even in early efficacy studies of prophylaxis, explanatory (per protocol) analysis may be misleading. Whilst it is unhelpful at this stage to include patients with random major protocol violations, drop-outs may be treatment-related. Analysis should therefore be based on the intention-to-treat (ITT) population. Since time to onset of effect is of interest, analysis of efficacy should be for the entire treatment period as well as for the period specified by the primary endpoint (e.g., the final 4 weeks of treatment). Standard statistical methods are appropriate. Adverse events are usually analysed descriptively.

II.3.ii. Chronic tension-type headache prophylaxis

II.3.ii.A. Objectives
To evaluate efficacy against chronic tension-type headache.

II.3.ii.B. Primary end-points
  a. Number of days with headache per specified unit time (usually 4 weeks) measured during treatment after a specified period (at least 8 weeks).
  b. Response rate: percentage of patients with reduction in headache days per unit time of 50% or more (implying reversion from chronic to episodic tension-type headache) after a specified treatment period.

II.3.ii.C. Secondary endpoints
  a. Number of days with headache over the entire treatment period.
  b. Intensity of headache on a visual analogue scale or 4-point verbal rating scale [0 = no pain; 1, 2, 3 = mild, moderate, severe pain]) averaged over attacks within a specified evaluation period.
  c. Duration of headache each day.
  d. Drug consumption for symptomatic or acute treatment totalled over an evaluation period.
  e. Incidence and nature of adverse events.

II.3.ii.D. Study design
These are invariably medium-term studies (at least 4 months) conducted in outpatients. Recommended are randomised, double-blind, placebo-controlled parallel-groups studies. There are no licensed active comparators. Randomisation should occur after a run-in (baseline) period of at least one month during which the number of days with headache and acute or symptomatic medication consumption are recorded. Stratification is unnecessary. Treatment periods should be at least 3 months. Days with headache, intensity and duration of headache, medication use and adverse events should be recorded as
they occur by patients in paper or electronic diaries. Reviews of patients and their diaries should be undertaken every 4 weeks. Acute medication is inappropriate treatment for this disorder and should not be encouraged (regular use of acute or symptomatic medication on >2 days per week will put the diagnosis in question as this approaches the threshold for medication-overuse headache).

Compliance with preventative medication has not been evaluated in chronic tension-type headache. It may be better than in migraine because symptoms are present daily or on most days rather than intermittently. Nevertheless, in phase II it is important to ascertain that the drug has been taken as prescribed. Consideration should be given to using electronic event monitors.

II.3.ii.E. Planned sample
Sample size calculations should be based on demonstrating superiority over placebo in a primary analysis of difference in days with headache. A change from baseline of ≥50% represents a clinically significant benefit of treatment, but the response to placebo has not been well documented (a reduction of up to 30% should be anticipated). Alternatively the primary analysis may be of difference in response rates. An absolute difference of 20% would be clinically significant. Again the response rate to placebo has not been well documented but up to 30% should be anticipated.

II.3.ii.F. Study population
Adults with chronic tension-type headache drawn from secondary or primary care or from the general population.

II.3.ii.G. Specific inclusion criteria
a. Patients with chronic tension-type headache conforming to IHS diagnostic criteria 2.3 for at least 3 months and with at least 3 months’ well-documented retrospective history.
b. Males and females.
c. Unless otherwise justified, patients should be over 18 years of age.

II.3.ii.H. Specific exclusion criteria
a. Age at onset of chronic tension-type headache of 50 years or over.
b. Other headaches, especially migraine, not well distinguished from tension-type headache or occurring with such frequency as to interfere with assessments.
c. Other illnesses, particularly depression, likely to interfere with assessments.
d. Use of other prophylactic drugs in the previous month.
e. Use of acute or symptomatic medication for headache on an average of >2 days per week over the previous 2 months.
f. Other history of drug or alcohol overuse.
g. Use of or requirement for psychotropic medication or other unacceptable concomitant therapy.
h. Risk of pregnancy.

II.3.ii.I. Tools for assessing endpoints
Paper or electronic diaries.

II.3.ii.J. Specific criteria for early withdrawal and discontinuation
a. Withdrawal of consent (which may be related to lack of efficacy or adverse events).
b. Medical concerns related to evident lack of efficacy or adverse events.
c. Other intercurrent illness.
d. Pregnancy.
e. Breach of double-blinding.
II.3.ii.K. Data analysis method
Even in early efficacy studies of prophylaxis, explanatory (per protocol) analysis may be misleading. Whilst it is unhelpful at this stage to include patients with random major protocol violations, drop-outs may be treatment-related. Analysis should therefore be based on the intention-to-treat (ITT) population. Since time to onset of effect is of interest, analysis of efficacy should be for the entire treatment period as well as for the period specified by the primary endpoint (e.g., the final 4 weeks of treatment). Standard statistical methods are appropriate. Adverse events are usually analysed descriptively.

II.3.iii. Prophylaxis of episodic cluster headache

II.3.iii.A. Objectives
To evaluate efficacy in terminating a cluster period or in reducing frequency, intensity and/or duration of continuing cluster headache attacks.

II.3.iii.B. Primary end-points
a. Frequency of attacks per specified unit time (usually 1 week) measured during treatment after a specified period (to allow treatment effect to develop) following dosage-stabilisation.

b. Remission rate: percentage of patients whose attacks have ceased after a specified treatment period.

The number of attacks should be recorded irrespective of their intensity or duration. An attack treated successfully with acute medication but with relapse within 1 hour counts as one attack.

II.3.iii.C. Secondary endpoints
a. Frequency of attacks over the entire treatment period.

b. Time to remission.

c. Intensity of cluster headaches averaged over a specified evaluation period.

d. Duration of cluster headaches summed or averaged over a specified evaluation period.

e. Drug consumption for symptomatic or acute treatment totalled over an evaluation period.

f. Incidence and nature of adverse events.

II.3.iii.D. Study design
In phase II these are short-term randomised, double-blind, placebo-controlled, parallel-groups studies conducted in outpatients. No run-in (baseline) period is needed. Stratification is recommended for time since onset of the cluster period (e.g., ≥2 or <2 weeks, but see below) and gender as each may influence the prophylactic effect or spontaneous remission rate. Treatment periods may be defined by the times prescribed for the primary end-point but should be at least 2 weeks; although they may need to incorporate dose-titration, they should not be substantially longer than this since treatments include placebo. Patients should take their usual acute therapy whenever cluster headache is of at least moderate intensity provided that it can be safely administered with the study drug. Attacks and their intensity and duration (and, if required, their associated features), acute medication use and adverse events should be recorded as they occur by patients in paper or electronic diaries. Reviews of patients and their diaries should be undertaken every week.

Compliance should be monitored. Because of the symptom frequency and severity it may be better in cluster headache than in other headache disorders but consideration should be given to using electronic event monitors.

II.3.iii.E. Planned sample
Sample size calculations should be based on demonstrating superiority over placebo in a primary analysis of a) difference in attack frequencies, with a relative difference of 50% being clinically significant and
allowing for a reduction on placebo of up to 20%; or b) difference in remission rates, with an absolute difference of 20% being clinically significant and allowing for a placebo plus spontaneous resolution rate of up to 20%.

II.3.iii.F. Study population
Adults with episodic cluster headache.

II.3.iii.G. Specific inclusion criteria
a. Patients with episodic cluster headache conforming to IHS diagnostic criteria 3.1, and in at least their second cluster period.
b. Expected duration of cluster period, from start of study medication, greater than the treatment period specified by the primary end-point (to limit the spontaneous-resolution rate in phase II, there may be advantage in restricting recruitment to patients within 2 weeks of onset of the cluster period).
c. Acute attacks occurring between once every 2 days and 5 times per day.
d. Males and females.
e. Unless otherwise justified, patients should be over 18 years of age.

II.3.iii.H. Specific exclusion criteria
a. Other headaches not well distinguished from cluster headache.
b. Other illnesses likely to interfere with assessments.
c. Other cluster headache prophylactic therapy in the previous week.
d. Use of or requirement for other unacceptable concomitant therapy.
e. Risk of pregnancy.
f. History of drug or alcohol overuse.

II.3.iii.I. Tools for assessing endpoints
Paper or electronic diaries.

II.3.iii.J. Specific criteria for early withdrawal and discontinuation
a. Withdrawal of consent (which may be related to lack of efficacy or adverse events).
b. Medical concerns related to evident lack of efficacy or adverse events.
c. Other intercurrent illness.
d. Pregnancy.
e. Breach of double-blinding.

II.3.iii.K. Data analysis method
Drop-outs may be treatment-related so, even in early efficacy studies of prophylaxis, analysis should be based on the intention-to-treat (ITT) population. Explanatory (per protocol) analysis may be worthwhile and hypothesis-generating as a secondary analysis. Subgroup analysis for gender differences is recommended and should be specified a priori. Since time to onset of effect is of interest, analysis of efficacy should be for the entire treatment period as well as for the period specified by the primary endpoint. Standard statistical methods are appropriate. Adverse events are usually analysed descriptively.

II.3.iv. Prophylaxis of chronic cluster headache
The objective in chronic cluster headache prophylaxis is long-term suppression of attacks. No methodology has yet been developed. In view of the difficulties with use of placebo as a comparator it is likely to be similar to that used in epilepsy, with add-on therapy studies preceding monotherapy trials.
III. PHASE III STUDIES FOR REGISTRATION OF NEW DRUGS

III.1. Outline of a typical development plan

During phase III, promising drugs are assessed for effectiveness against placebo in at least two pivotal large multicentre studies. They will be randomised double-blind trials incorporating one or more doses of study drug according to the findings of phase II. They may use parallel-groups or (multiple) cross-over designs, with regulators strongly favouring the former. For acute and prophylactic migraine therapy, and acute treatment of episodic tension-type headache and cluster headache, one or both of these, or additional studies in this phase, should include active comparators.

Acute treatments for cluster headache, which are unlikely to be oral, must be formulated so that patients can self-medicate wherever they may be at time of onset.

Pivotal acute-treatment trials should address not only the treatment of single attacks but also consistency of the therapeutic response across multiple attacks. Additionally in migraine and episodic tension-type headache, long-term continuation protocols are desirable to demonstrate repeatability of effect over time (lack of tachyphylaxis). Such studies contribute helpfully to safety evaluation.

In migraine particularly, the headache is not a stable pain but develops gradually, or sometimes rapidly, to a peak with subsequent spontaneous resolution. This poses challenges regarding timing of intake of medication. Trials in phase III may explore the relationship between timing of acute treatment and effect. Such a study in migraine with aura may incorporate medicating during the aura phase.

Specific trials in migraine and episodic tension-type headache may look at re-medicating, within the same attack, with a second dose of study drug when the first has been inadequately efficacious.

Prophylactic trials require a minimum of three months’ treatment for migraine or chronic tension-type headache, but are better designed to reflect treatment periods likely in routine management, which are typically longer (4-6 months or more). Furthermore, continued observation beyond the treatment period, for continuing efficacy or possibly rebound exacerbation, is essential.

Cluster headache to some extent has the status of orphan disease. Long-term prophylaxis may be inappropriate in the episodic subtype, when attacks recur over periods of only a few weeks. Some currently used drugs achieve remission quite quickly, even within a few days, and prolonged treatment may not be necessary. This, and the fact that no treatments currently used for the prevention of cluster headache are licensed for this indication, make active-comparator studies difficult whilst placebo cannot be used long-term. The regulatory requirements for phase III have not been clarified.

In all primary headache disorders, safety of treatment is a major concern since the disorders themselves are self-limiting. On the other hand they are widespread and drugs that treat them, once marketed, are likely to be taken by large numbers of people and not always in strict accordance with instructions. Regulators will look carefully at safety, and may require special studies in diseased populations.

Headache sufferers attending specialist clinics may not be representative of the larger number seen by primary-care physicians. Neither group is likely to match those in the general population who do not seek medical advice. Phase III trials need to recruit widely from the population who will use the agent when marketed, with as few restrictions as possible. Nonetheless, special protocols will be required for children (under the age of 18), whose needs may be different, and may be required for the elderly (over the age of 65), who are less subject to primary headaches and more at risk of symptomatic headache as well as other illness.
III.2. Short term studies

III.2.i. Acute treatment of migraine

III.2.i.A. Objectives
To confirm effectiveness and evaluate comparative efficacy and tolerability in relieving symptoms of the acute attack.

III.2.i.B. Primary end-point
a. “Pain-free” rate: percentage of patients pain-free at 2 hours after treatment.

III.2.i.C. Secondary endpoints
a. Percentage of patients pain-free at 2 hours and remaining pain-free, without use of further medication, for 48 hours after treatment.
b. Rate of relapse, defined as the return of headache of any intensity within 48 hours in patients pain-free at 2 hours after treatment.
c. Headache intensity at various time points after treatment.
d. “Headache-relief” rate: percentage of patients with a decrease in headache intensity from severe or moderate to mild or no pain within 2 hours after treatment.
e. Time to “meaningful” pain relief (usually defined subjectively).
f. Time to “onset of action” (defined as first noticeable pain relief).
g. Functional disability on a validated scale (usually a 4-point verbal rating scale where 0 = no functional impairment, 1 = “can do everything albeit with difficulty”, 2 = “cannot do some things” and 3 = “cannot do anything and/or bed-bound”) at 2 hours and other time points after treatment.
h. Effect on associated symptoms such as nausea, vomiting, photophobia and phonophobia.
i. Rate and timing of use of rescue medication.
j. Global evaluation of study medication.
k. Pharmacoeconomic measures.
l. Incidence and nature of adverse events.

III.2.i.D. Study design
Short-term studies should be randomised, double-blind, placebo-controlled parallel-groups outpatient trials treating one attack per patient. Three-arm trials, including placebo, are required for internal validation in active-comparator studies because of the large and highly variable placebo effect in migraine studies. In general, there is no need for stratification but phase III studies of acute treatment may opt to include patients with or without specific prophylactic medication(s), in which case stratification is based on this variable. Outcome variables are usually recorded by the patient in paper or electronic diaries, with prompts at various time points. Rescue medication should be allowed after 2 hours. This may be replaced, in patients who first received active drug, by a second dose of study drug or placebo in re-medication trials, with rescue medication deferred to 4 hours. To maintain the double-blind without subjecting patients to placebo only for 4 hours, those who first received placebo are given active drug as the second dose. The observation period after treatment of an attack should be 48 hours. Patients should return for final review soon after this.

III.2.i.E. Planned sample
Sample size calculations should be based on demonstrating superiority over placebo in a primary analysis of difference in pain-free rates, with an absolute difference of 20% being clinically significant and allowing for a placebo rate of up to 20%.

In practice, much greater numbers are required to demonstrate safety. Regulators require data from a large and representative group of patients. Trials including >1,000 patients are not unusual.
III.2.i.F. Study population
   a. Adults with migraine with or without aura.
   b. Adolescents and/or children, if they are to be included in the labelling.

III.2.i.G. Specific inclusion criteria
   a. Patients with migraine conforming to IHS diagnostic criteria 1.1 or 1.2 for at least 1 year and with at least 3 months’ well-documented retrospective history.
   b. Migraine attacks occurring 1-6 times monthly.
   c. Males and females.

At the time of treatment:
   a. An acute attack, usually with onset within the previous 12 hours.
   b. At least 48 hours since resolution of the previous attack.
   c. Headache of moderate or severe intensity.
   d. So far untreated.

III.2.i.H. Specific exclusion criteria
   a. Age at onset of migraine of 50 years or over.
   b. Other headaches not well distinguished from migraine or occurring with such frequency as to interfere with assessments (usually taken to mean occurring on >6 days/month).
   c. Other illnesses likely to interfere with assessments.
   d. Use of migraine prophylactic drugs in the previous month or, if the protocol allows inclusion of patients on prophylaxis, any change in nature or dose of prophylactic medication in the previous 3 months.
   e. Use of or requirement for other unacceptable concomitant therapy.
   f. History of drug or alcohol overuse.

III.2.i.I. Tools for assessing endpoints
Paper or electronic diaries, with prompts at various time points.

III.2.i.J. Data analysis method
Analysis should be based on the intention-to-treat (ITT) population, although this may be defined to exclude those known not to have taken treatment. Standard statistical methods are appropriate. Adverse events are usually analysed descriptively.

Cost-effectiveness (or cost-utility) analysis is highly desirable, but the methodology is not yet well developed.

III.2.ii. Acute treatment of episodic tension-type headache

III.2.ii.A. Objectives
To confirm effectiveness, and evaluate comparative efficacy and tolerability, in relieving pain and functional impairment attributable to acute episodic tension-type headache.

III.2.ii.B. Primary end-point
“Pain-free” rate: percentage of patients pain-free at 2 hours after treatment.

III.2.ii.C. Secondary endpoints
   a. Headache intensity (scored on either a visual analogue scale or a 4-point verbal rating scale [0 = no pain; 1, 2, 3 = mild, moderate, severe pain]) at 2 hours and other time points after treatment.
b. Headache intensity difference (the arithmetic change in headache intensity score) at 2 hours and at other time points after treatment.
c. Headache relief (on a verbal rating scale from “none” to “complete”, with two or more intermediaries which may include “meaningful relief”; negative scores may be incorporated to indicate worsening) at 2 hours and at other time points after treatment.
d. Functional disability on a validated scale (e.g., a 4-point verbal rating scale where 0 = no functional impairment, 1 = “can do everything albeit with difficulty”, 2 = “cannot do some things” and 3 = “cannot do anything and/or bed-bound”) at 2 hours and other time points after treatment.
e. Rate and timing of use of rescue medication.
f. Global evaluation of study medication.
g. Incidence and nature of adverse events.

III.2.ii.D. Study design
Short-term studies should be randomised, double-blind, placebo-controlled parallel-groups outpatient trials treating one attack per patient. Three-arm trials, including placebo, are required for internal validation in active-comparator studies because of the very large placebo effect reported in acute episodic tension-type headache studies. There is no need for stratification. Outcome variables are usually recorded by the patient in paper or electronic diaries, with prompts at various time points. Rescue medication should be allowed after 2 hours. This may be replaced, in patients who first received active drug, by a second dose of study drug or placebo in re-medication trials, with rescue medication deferred to 4 hours. To maintain the double-blind without subjecting patients to placebo only for 4 hours, those who first received placebo are given active drug as the second dose. The observation period after treatment should be at least 24 hours. Patients should return for final review soon after this.

III.2.ii.E. Planned sample
Sample size calculations should be based on demonstrating superiority over placebo in a primary analysis of difference in pain-free rates, with an absolute difference of 20% being clinically significant and allowing for a placebo rate of up to 50%.

If the study drug is not already licensed for another indication, much greater numbers may be required to demonstrate safety.

III.2.ii.F. Study population
a. Adults with episodic tension-type headache drawn from the general population (by advertising if necessary).
b. Adolescents and/or children, if they are to be included in the labelling.

III.2.ii.G. Specific inclusion criteria
a. Patients with frequent episodic tension-type headache (occurring on >1 but <15 days per month) conforming to IHS diagnostic criteria 2.2 for at least 1 year and with at least 3 months’ well-documented retrospective history.
b. Usual headache duration at least 4 hours.
c. Males and females.

At the time of treatment:
a. An acute episode of tension-type headache, usually with onset within the previous 12 hours.
b. Headache of at least moderate intensity.
c. So far untreated.

III.2.ii.H. Specific exclusion criteria
a. Age at onset of tension-type headache of 50 years or over.
b. Chronic tension-type headache.
c. Migraine if not well distinguished from tension-type headache or occurring in the previous year more frequently than once per month.
d. Medication-overuse headache.
e. Other headaches not well distinguished from tension-type headache or occurring with such frequency as to interfere with assessments.
f. Other illnesses likely to interfere with assessments.
g. Use of prophylactic drugs in the previous month.
h. Use of or requirement for psychotropic medication or other unacceptable concomitant therapy.
i. History of drug or alcohol overuse.

III.2.ii.1. Tools for assessing endpoints
Paper or electronic diaries, with prompts at various time points.

III.2.ii.J. Data analysis method
Analysis should be based on the intention-to-treat (ITT) population, although this may be defined to exclude those known not to have taken treatment. Standard statistical methods are appropriate. Adverse events are usually analysed descriptively.

Longer-term studies should address consistency of the therapeutic response across attacks. These may be double-blind cross-over studies observing treatment of several attacks per patient with the same drug and dose plus, randomly, one or more (e.g., one attack out of five) with placebo. Additionally, long-term continuation protocols are desirable to demonstrate repeatability of effect over time (lack of tachyphylaxis). Such studies contribute helpfully to safety evaluation. They need be neither placebo-controlled nor blinded.

III.2.iii. Acute treatment of episodic or chronic cluster headache

III.2.iii.A. Objectives
To evaluate efficacy and comparative effectiveness and tolerability in aborting or suppressing the acute attack.

III.2.iii.B. Primary end-points
a. “Aborted attack” rate: percentage of patients in whom the attack is effectively stopped (headache intensity reduced to mild or no pain) within a prescribed time interval (which may be as short as 10 minutes).
b. Time to meaningful relief.
c. Time to “complete” relief (mild or no pain).

III.2.iii.C. Secondary endpoints
a. Rate of relapse, defined as the return of headache of moderate or greater intensity within 1 hour in patients reporting an aborted attack.
b. Headache intensity (on a 5-point verbal rating scale: 0 = no pain, 1, 2, 3, 4 = mild, moderate, severe, excruciating pain) at 5, 10 and 15 minutes after treatment and every 15 minutes thereafter for up to 3 hours (whilst these repeated assessments are recommended, marked agitation is a feature of acute cluster headache which, combined with severe pain, may make them impractical).
c. Effect on associated autonomic symptoms.
d. Functional impairment on a validated scale.
e. Rate and timing of use of rescue medication.
f. Global evaluation of study medication.
g. Patient’s preference (in cross-over studies).
h. Incidence and nature of adverse events.
III.2.iii.D. Study design
Treatments coming into phase III may be oral but are more likely to be parenteral. There are study-design implications for parenteral therapy, particularly for active-comparator studies and especially because the active comparator may itself be administered parenterally.

Short-term studies should be randomised, double-blind, placebo-controlled parallel-groups or cross-over trials in outpatients treating 1-4 attacks each. The cross-over design has advantages, and may be accepted by regulators, since patients are uncommon whilst attacks occur with high and predictable frequency. Three-arm trials, including placebo, are required for internal validation in active-comparator studies; although placebo effect is relatively slight in cluster headache, trials are easily confounded by high rates of spontaneous resolution of attacks which are short-lasting. Active-comparator studies are likely to require a double-dummy design if one or other treatment, or both, is administered parenterally. At least one large multicentre trial should formally confirm efficacy.

If a parallel-groups trial includes both episodic and chronic cluster headache, stratification is recommended because responses to treatment may differ. Stratification for gender is also recommended for the same reason. Outcome variables are usually recorded by the patient in paper or electronic diaries, with prompts at various time points. Rescue medication should be allowed after the time prescribed for the primary end-point, but options for this are very limited. The observation period after treatment of an attack should be not less than 24 hours unless interrupted by the occurrence of the next attack. Depending on the experience from phase II, reviews may shortly follow each treatment or (in multiple-attack studies) only the last treatment.

Short-term studies in cluster headache can address consistency of the therapeutic response across attacks. These may be double-blind cross-over studies, observing treatment of several attacks per patient with the same drug and dose plus, randomly, one or more (e.g., one attack out of five) with placebo.

III.2.iii.E. Planned sample
Sample size calculations should be based on demonstrating superiority over placebo in a primary analysis of difference in aborted-attack rates. Placebo-response rate is low but, because of the short attack-duration, spontaneous remission rates may be high; the two may combine to 50%. An absolute difference of 25% is clinically significant.

III.2.iii.F. Study population
a. Adults with episodic or chronic cluster headache.
   b. Adolescents and/or children, if they are to be included in the labelling.

III.2.iii.G. Specific inclusion criteria
a. Patients with episodic or chronic cluster headache conforming to IHS diagnostic criteria 3.1 or 3.2; patients with episodic cluster headache should be in at least their second cluster period.
   b. Acute attacks occurring between once every 2 days and 5 times per day.
   c. Attack duration of 30-180 minutes.
   d. Males and females.

At the time of treatment:
   a. An acute attack, usually with onset within the previous 15 minutes (at least 15 minutes before expected spontaneous resolution).
   b. At least 1 hour since resolution of the previous attack and 24 hours (or 5 half-lives if longer) since the latest previous use of study drug.
   c. Headache of moderate or greater intensity.
   d. So far untreated.
III.2.iii.H. Specific exclusion criteria
   a. Other headaches not well distinguished from cluster headache.
   b. Other illnesses likely to interfere with assessments.
   c. Use of or requirement for unacceptable concomitant therapy.
   d. History of drug or alcohol overuse.

III.2.iii.I. Tools for assessing endpoints
Paper or electronic diaries, with prompts at various time points.

III.2.iii.J. Data analysis method
Analysis should be based on the intention-to-treat (ITT) population, although this may be defined to exclude those known not to have taken treatment. Subgroup analyses (for episodic and chronic subtypes and for gender differences) are recommended and should be specified a priori. Standard statistical methods are appropriate. “Time to” end-points require survival-analysis methods. Adverse events are usually analysed descriptively.

The frequency of medication in cluster headache may lead to high treatment costs; given that, untreated, this disorder is very disabling and ruins quality of life, cost-effectiveness (or cost-utility) analysis is appropriate but the methodology is not yet well developed.

Longer-term studies are desirable, extending throughout a cluster episode or in patients with chronic cluster headache, to demonstrate repeatability of effect over time (lack of tachyphylaxis) and for safety evaluation. These need be neither placebo-controlled nor blinded.

III.3. Long term studies

Longer-term studies should address consistency of the therapeutic response across attacks. These may be double-blind cross-over studies observing treatment of several attacks per patient with the same drug and dose plus, randomly, one or more (e.g., one attack out of five) with placebo. Additionally, long-term continuation protocols are desirable to demonstrate repeatability of effect over time (lack of tachyphylaxis). At least one study of 12 months’ duration is needed for safety evaluation. These need be neither placebo-controlled nor blinded.

One or more of these protocols should accommodate the double-blind investigation, using similar end-points, of re-medication to treat relapse following initial successful treatment with the study drug.

III.3.i. Migraine prophylaxis

III.3.i.A. Objectives
To confirm effectiveness and evaluate comparative efficacy and tolerability in migraine prevention.

III.3.i.B. Primary end-points
   a. Frequency of attacks per specified unit time (usually 4 weeks) measured during treatment after a specified period (usually 8 weeks).
   b. Response rate: percentage of patients with frequency reduction of 50% or more after a specified treatment period.

The number of attacks should be recorded irrespective of their duration, and the following rules distinguish an attack of long duration from two attacks and between attacks and relapses:
a. A migraine attack which is interrupted by sleep, or which temporarily remits spontaneously and then recurs within 48 hours after its onset, should be recorded as one attack and not two.
b. An attack treated successfully with medication but with relapse within 48 hours counts as one attack.

III.3.i.C. Secondary endpoints
a. Frequency of attacks over the entire treatment period.
b. Frequency of attacks following discontinuation of treatment.
c. “Migraine days” (defined as any day on which symptoms of migraine are present) per 4 weeks.
d. Intensity of migraine headache averaged over attacks within a specified evaluation period.
e. Speed of effect (e.g., response rates in first, second and third months of treatment).
f. Drug consumption for symptomatic or acute treatment totalled over an evaluation period.
g. Headache indices multiplying frequency, intensity and/or duration are not recommended: arbitrary weighting in the numerical scores, which may be faulty, is increased by multiplication; and indices cannot meaningfully be compared between patients.
h. Health-related quality of life measures would be desirable but none are well-established or universally accepted; they should not be used until clinically validated.
i. Global evaluation of study medication.
j. Pharmacoeconomic measures.
k. Incidence and nature of adverse events.

III.3.i.D. Study design
These are medium- or long-term studies (at least 4 months) in outpatients. At least two should be randomised, double-blind, placebo-controlled parallel-groups studies and at least one of these should include an active comparator. Three-arm trials, including placebo, are required for internal validation with active comparators unless the study is designed to show superiority over a well-established comparator (if superiority is not shown, non-inferiority cannot be claimed in the absence of placebo control). Randomisation should occur after a run-in (baseline) period of at least one month, when stratification for baseline attack rate (e.g., ≥3 or <3 per 4 weeks) is recommended as the prophylactic effect may depend on this variable. Treatment periods should be at least 3 months. Patients should take their usual acute therapy as required provided that it can be safely administered with the study drug. Attacks (and, if required, their features), acute medication use and adverse events should be recorded as they occur by patients in paper or electronic diaries. Reviews of patients and their diaries should be undertaken every 4 weeks, and should continue for at least 4 weeks after treatment is discontinued.

Compliance should be monitored.

III.3.i.E. Planned sample
Sample size calculations should be based on demonstrating superiority over placebo in a primary analysis of a) difference in attack frequencies, with a relative difference of 50% or an absolute difference of 1 attack/month being clinically significant and allowing for a reduction on placebo of up to 30% or 1 attack/month; or b) difference in responder rates, with an absolute difference of 20% being clinically significant and allowing for a placebo rate of up to 30%.

Whether or not greater numbers are required for safety evaluation is dependent on whether or not the study drug has been through a development programme, and is licensed already, for another indication.

III.3.i.F. Study population
a. Adults with frequent attacks of migraine with or without aura.
b. Adolescents and/or children, if they are to be included in the labelling.
III.3.i.G. Specific inclusion criteria
a. Patients with migraine conforming to IHS diagnostic criteria 1.1 or 1.2 for at least 1 year and with at least 3 months’ well-documented retrospective history.
b. Migraine attacks occurring 2-6 times monthly.
c. Males and females.

III.3.i.H. Specific exclusion criteria
a. Age at onset of migraine of 50 years or over.
b. Other headaches not well distinguished from migraine or occurring with such frequency as to interfere with assessments (usually taken to mean occurring on >6 days/month).
c. Other illnesses likely to interfere with assessments.
d. Use of other migraine prophylactic drugs in the previous month.
e. Use of or requirement for other unacceptable concomitant therapy.
f. Risk of pregnancy.
g. History of drug or alcohol overuse.

III.3.i.I. Tools for assessing endpoints
Paper or electronic diaries.

III.3.i.J. Specific criteria for early withdrawal and discontinuation
a. Withdrawal of consent (which may be related to lack of efficacy or adverse events).
b. Medical concerns related to evident lack of efficacy or adverse events.
c. Other intercurrent illness.
d. Pregnancy.
e. Breach of double-blinding.

III.3.i.K. Data analysis method
Analysis should be based on the intention-to-treat (ITT) population. Because time to onset of effect is of interest, analysis of efficacy should be for the entire treatment period as well as for the period specified by the primary endpoint (e.g., the final 4 weeks of treatment). Standard statistical methods are appropriate. Adverse events are usually analysed descriptively.

Cost-effectiveness (or cost-utility) analysis is highly desirable, but the methodology is not yet well developed.

Longer-term studies are desirable to investigate continuing efficacy during and after periods of treatment (up to 6 months or longer) common in routine practice. At least one trial of at least 12 months’ duration is required for safety evaluation. These studies can be conducted as continuations of double-blind studies with patients opting, or not, to remain on their treatment (whether active or placebo). In addition, withdrawal of medication at the end of the prescribed period of treatment should be evaluated, ideally by randomised and double blind substitution of placebo in one group of patients and continuation of active therapy in another (with informed consent). Open observational studies, using patients as their own controls, have very limited value and are a poor alternative because of the inherent variability over time of the disease.

III.3.ii. Chronic tension-type headache prophylaxis

III.3.ii.A. Objectives
To confirm efficacy, effectiveness and tolerability in treating chronic tension-type headache.

III.3.ii.B. Primary end-points
a. Number of days with headache per specified unit time (usually 4 weeks) measured during treatment after a specified period (at least 8 weeks).
b. Response rate: percentage of patients with reduction in headache days per unit time of 50% or more (implying reversion from chronic to episodic tension-type headache) after a specified treatment period.

III.3.ii.C. Secondary endpoints

a. Number of days with headache over the entire treatment period.
b. Intensity of headache on a visual analogue scale or 4-point verbal rating scale \([0 = \text{no pain}; \ 1,\ 2,\ 3 = \text{mild, moderate, severe pain}]\) averaged over attacks within a specified evaluation period.
c. Duration of headache each day.
d. Headache indices multiplying frequency, intensity and/or duration are not recommended: arbitrary weighting in the numerical scores, which may be faulty, is increased by multiplication; and indices cannot meaningfully be compared between patients.
e. Functional measures and health-related quality of life measures would be desirable but are not established and should not be used until clinically validated.
f. Drug consumption for symptomatic or acute treatment totalled over an evaluation period.
g. Global evaluation of study medication.
h. Incidence and nature of adverse events.

III.3.ii.D. Study design

These are medium- or long-term studies (at least 4 months) in outpatients. At least two should be randomised, double-blind, placebo-controlled parallel-groups studies. There are no licensed active comparators. Randomisation should occur after a run-in (baseline) period of at least one month during which the number of days with headache and acute or symptomatic medication consumption are recorded. Stratification is unnecessary. Treatment periods should be at least 3 months. Days with headache, intensity and duration of headache, acute medication use and adverse events should be recorded as they occur by patients in paper or electronic diaries. Reviews of patients and their diaries should be undertaken every 4 weeks, and should continue for at least 4 weeks after treatment is discontinued.

Acute medication is inappropriate treatment for this disorder and should not be encouraged (regular use of acute or symptomatic medication on >2 days per week will put the diagnosis in question as this approaches the threshold for medication-overuse headache). Compliance should be monitored.

III.3.ii.E. Planned sample

Sample size calculations should be based on demonstrating superiority over placebo in a primary analysis of difference in days with headache. A change from baseline of \(\geq 50\%\) represents a clinically significant benefit of treatment, but the response to placebo has not been well documented (a reduction of up to 30\% should be anticipated). Alternatively the primary analysis may be of difference in response rates. An absolute difference of 20\% would be clinically significant. Again the response rate to placebo has not been well documented but up to 30\% should be anticipated.

III.3.ii.F. Study population

Adults with chronic tension-type headache drawn from secondary or primary care or from the general population.

III.3.ii.G. Specific inclusion criteria

a. Patients with chronic tension-type headache conforming to IHS diagnostic criteria 2.3 for at least 3 months and with at least 3 months’ well-documented retrospective history.
b. Males and females.
c. Unless otherwise justified, patients should be over 18 years of age.
III.3.ii.H. Specific exclusion criteria

a. Age at onset of chronic tension-type headache of 50 years or over.
b. Other headaches, especially migraine, not well distinguished from tension-type headache or occurring with such frequency as to interfere with assessments.
c. Other illnesses, particularly depression, likely to interfere with assessments.
d. Use of other prophylactic drugs in the previous month.
e. Use of acute or symptomatic medication for headache on an average of >2 days per week over the previous 2 months.
f. Other history of drug or alcohol overuse.
g. Use of or requirement for psychotropic medication or other unacceptable concomitant therapy.
h. Risk of pregnancy.

III.3.ii.I. Tools for assessing endpoints

Paper or electronic diaries.

III.3.ii.J. Specific criteria for early withdrawal and discontinuation

a. Withdrawal of consent (which may be related to lack of efficacy or adverse events).
b. Medical concerns related to evident lack of efficacy or adverse events.
c. Other intercurrent illness.
d. Pregnancy.
e. Breach of double-blinding.

III.3.ii.K. Data analysis method

Analysis should be based on the intention-to-treat (ITT) population. Because time to onset of effect is of interest, analysis of efficacy should be for the entire treatment period as well as for the period specified by the primary endpoint (e.g., the final 4 weeks of treatment). Standard statistical methods are appropriate. Adverse events are usually analysed descriptively.

Longer-term studies are desirable to investigate continuing efficacy over longer periods of treatment that may be necessary in routine management (6 months or longer). At least one trial of at least 12 months’ duration is required for safety evaluation. These studies can be conducted as continuations of double-blind studies with patients opting, or not, to remain on their treatment (whether active or placebo). In addition, withdrawal of medication at the end of the prescribed period of treatment should be evaluated, ideally by randomised and double-blind substitution of placebo in one group of patients and continuation of active therapy in another (with informed consent). Open observational studies, using patients as their own controls, are of very limited value and a poor alternative because of the inherent variability over time of the disease.

III.3.iii. Prophylaxis of episodic cluster headache

III.3.iii.A. Objectives

To confirm efficacy and evaluate effectiveness and tolerability in terminating a cluster period or in reducing frequency, intensity and/or duration of continuing cluster headache attacks.

III.3.iii.B. Primary end-points

a. Frequency of attacks per specified unit time (usually 1 week) measured during treatment after a specified period (to allow treatment effect to develop) following dosage-stabilisation.
b. Remission rate: percentage of patients whose attacks have ceased after a specified treatment period.
The number of attacks should be recorded irrespective of their intensity or duration. An attack treated successfully with acute medication but with relapse within 1 hour counts as one attack.

III.3.iii.C. Secondary endpoints
   a. Frequency of attacks over the entire treatment period.
   b. Time to remission.
   c. Intensity of cluster headaches averaged over a specified evaluation period.
   d. Duration of cluster headaches summed or averaged over a specified evaluation period.
   e. Drug consumption for symptomatic or acute treatment totalled over an evaluation period.
   f. Health-related quality of life measures, and measures of functional disability over the treatment period, would be desirable but are not established and should not be used until clinically validated.
   g. Global evaluation of study medication.
   h. Effects after withdrawal of treatment.
   i. Incidence and nature of adverse events.

III.3.iii.D. Study design
These are short- or medium-term studies depending on the treatment effect observed in phase II. However, studies over >2 weeks cannot be conducted against placebo notwithstanding that high spontaneous remission rates confound trials that are not placebo-controlled. Therefore, superiority over an established comparator must be shown in one or more randomised parallel-groups studies. Whilst a number of reasonably effective potential comparator drugs exist, they are unlicensed for this indication, associated with toxicity and tend to be used in ways that make it very difficult to achieve double-blindness. Open studies are more acceptable with objective end-points (e.g., remission rate).

No run-in (baseline) period is needed. Stratification is recommended for time since onset of the cluster period (e.g., ≥2 or <2 weeks) and gender as each may influence the prophylactic effect or spontaneous remission rate. Treatment periods may need to incorporate dose-titration and, following dosage stabilisation, are defined by the times prescribed for the primary end-point or by the study objective if this calls for longer-term therapy. They are unlikely to exceed 3 months and safety evaluation must be conducted within this period unless safety has been demonstrated already in longer-term use of the drug for other indications.

Patients should take their usual acute therapy whenever cluster headache is of at least moderate intensity provided that it can be safely administered with the study drug. Attacks and their intensity and duration (and, if required, their associated features), acute medication use and adverse events should be recorded as they occur by patients in paper or electronic diaries. Reviews of patients and their diaries should be undertaken every week in short-term studies and at least monthly in medium-term studies.

Compliance should be monitored.

After remission of the cluster period, whether spontaneous or treatment-related, prophylactic medication is withdrawn. At least one trial should observe the consequences of withdrawal over up to several weeks, since these may include relapse.

III.3.iii.E. Planned sample
Sample size calculations should be based on demonstrating superiority over placebo in a primary analysis of a) difference in attack frequencies, with a relative difference of 50% being clinically significant and allowing for a reduction on placebo of up to 20%; or b) difference in remission rates, with an absolute difference of 20% being clinically significant and allowing for a placebo plus spontaneous resolution rate of up to 20%.
III.3.iii.F. Study population
   a. Adults with episodic cluster headache.
   b. Adolescents and/or children, if they are to be included in the labelling.

III.3.iii.G. Specific inclusion criteria
   a. Patients with episodic cluster headache conforming to IHS diagnostic criteria 3.1 and in at least
      their second cluster period.
   b. Any length of time from onset of the cluster period provided that its expected duration, from start of
      study medication, is greater than the treatment period specified by the primary end-point.
   c. Acute attacks occurring between once every 2 days and 5 times per day.
   d. Males and females.

III.3.iii.H. Specific exclusion criteria
   a. Other headaches not well distinguished from cluster headache.
   b. Other illnesses likely to interfere with assessments.
   c. Other cluster headache prophylactic therapy in the previous week.
   d. Use of or requirement for other unacceptable concomitant therapy.
   e. Risk of pregnancy.
   f. History of drug or alcohol overuse.

III.3.iii.I. Tools for assessing endpoints
   Paper or electronic diaries.

III.3.iii.J. Specific criteria for early withdrawal and discontinuation
   a. Withdrawal of consent (which may be related to lack of efficacy or adverse events).
   b. Medical concerns related to evident lack of efficacy or adverse events.
   c. Other intercurrent illness.
   d. Pregnancy.

III.3.iii.K. Data analysis method
   Analysis should be based on the intention-to-treat (ITT) population. Subgroup analysis for gender differences is
   recommended and should be specified a priori. Since time to onset of effect is of interest, analysis of efficacy
   should be for the entire treatment period as well as for the period specified by the primary endpoint. Standard
   statistical methods are appropriate. Adverse events are usually analysed descriptively.

IV. OTHER STUDIES (SPECIAL INDICATIONS AND PRAGMATIC STUDIES)

IV.1. Children and adolescents

   Development of drugs for headache disorders in these age-groups is clearly required for two indications:
   a. acute treatment of migraine;
   b. prophylaxis of migraine.

   Disease characteristics differ in migraine between children/adolescents and adults. The results of acute
   and prophylactic treatment studies in adults with migraine cannot be extrapolated to younger age-groups.
   Separate trials in children/adolescents are not required by regulators for initial marketing authorisation,
   but these age-groups will be excluded from product-labelling if sufficient efficacy and safety data are not
   included in the regulatory submission. It is not certain whether such differences exist in episodic and
   chronic tension-type headache, whilst cluster headache is very rare (although not unknown) in children. It
   is likely, however, that regulators will adopt the same approach in these disorders whilst markets may not
   be commercially viable.
Further dose-finding studies in these age-groups may be needed. During phase III, drugs with clear efficacy and safety in adults with migraine may be assessed in separate placebo-controlled trials for effectiveness and safety in children and/or adolescents. Pivotal studies will be large multicentre randomised double-blind trials incorporating one or more doses of study drug. They may use parallel-groups or (multiple) cross-over designs, with regulators strongly favouring the former. One or more studies should include an active comparator where licensed comparators exist. Three-arm trials, including placebo, are required for internal validation with active comparators unless the study is designed to show superiority over a well-established comparator (if superiority is not shown, non-inferiority cannot be claimed in the absence of placebo control).

Objectives, end-points, study designs, sample sizes, inclusion/exclusion criteria (other than age), tools for assessing end-points and data analysis methods are all generally similar to those in adult migraine trials. There are a few exceptions:

a. Migraine attacks are usually shorter-lasting, so rapid efficacy is more important;

b. Associated symptoms of nausea and vomiting are commonly more pronounced, and effective treatment of these may be a higher priority;

c. Children with headache are likely to be put to bed to sleep (which is curative), so frequent assessments over several hours is often impractical;

d. Prophylactic medication in children is inappropriate before a review has been conducted of lifestyle and possible triggers, which should be built into the protocol as part of baseline evaluation.

**IV.2. The elderly**

All primary headache disorders become significantly less prevalent after the age of 60 years. There are no special requirements for trials in the elderly, who should not generally be excluded from adult trials (subject to other inclusion/exclusion criteria). Elderly patients with migraine should have been suffering from this disorder for many years: onset of migraine over the age of 50 years is uncommon and predictive of symptomatic disease, which must be excluded.

**IV.3. Menstrual migraine**

Migraine in women may be hormonally-triggered and occur solely in close temporal relationship to menstrual periods (menstrual migraine) or it may be more loosely associated with menstruation with a tendency to occur at or around the time of periods (menstrually-associated migraine). It is unlikely that treatment of menstrually-associated migraine should differ from that of migraine generally, whereas other possibilities arise for the treatment of menstrual migraine.

The EMEA advises that studies in menstrual migraine, to be undertaken once efficacy and safety have been demonstrated in non-menstrual migraine, have in principle the same design and end-points as studies in non-menstrual migraine. Subgroups of patients with menstrual migraine included in several studies may be combined in a meta-analysis planned *a priori*. The temporal relationship between menses and migraine attacks should be stringently recorded, and for diagnostic purposes this is necessary for three cycles before trial entry. In acute treatment trials, an important secondary end-point is the percentage of patients pain-free at 2 hours and remaining pain-free, without use of further medication, for 48 hours after treatment. In prophylactic trials there is the option for monthly short-term perimenstrual prophylactic treatment. If this is being evaluated, the trial design should require continuous observation throughout each month since the possibility exists that attacks are merely postponed to later in the menstrual cycle.
IV.4. “Mild” migraine
The previously widely-adopted primary end-point for acute migraine treatment trials (“headache relief”) required that treatment was delayed until pain was moderate or severe. This is counter-intuitive and possibly counter-productive, and many patients will not do it routinely. Although it is in part justified by the argument that earlier treatment results in the inappropriate use of migraine-specific therapy for non-migraine headache, particularly episodic tension-type headache, this argument is not clearly evidence-based. There are good reasons for conducting additional trials of early treatment, whilst pain is still mild. The recommended primary end-point, “pain-free” rate (percentage of patients pain-free at 2 hours after treatment), can and should still be used. Secondary endpoints should include rate of relapse and percentage of patients pain-free at 2 hours and remaining pain-free, without use of further medication, for 48 hours after treatment.

IV.5. Acute treatment of migraine in the aura phase
Patients with migraine with aura have the opportunity to treat during the aura phase, before headache commences. Efficacy of treatment in this phase cannot be assumed from studies of treatment taken later in the attack. Separate studies are required, with similar endpoints.

IV.6. Pragmatic studies
Pragmatic trials attempt to replicate routine practice in the use of a drug rather than the conditions of a controlled experiment. They rarely support marketing authorisation applications but can usefully inform prescribing practitioners.

A major concern in acute migraine trials is that recommended end-points, chosen because they are relatively objectively measurable and have proved statistically robust in differentiating between active treatments and placebo, do not well reflect patients’ views of what they want from a treatment. One suggested design for a preference study dispenses to each patient a quantity of each of two or more comparator drugs (which, if blinded, can include placebo). The patient chooses which to use on the basis of accumulating personal experience of each. The rate of use of each is an index of preference. Other measures of “satisfaction” are needed also since preference for one treatment over another does not indicate that either is adequate.

No studies have yet compared acute migraine therapy alone with acute plus prophylactic therapy, but these are needed. End-points are likely to reflect quality-of-life or pharmacoeconomic measures.

V. STUDIES FOR THE REGISTRATION OF GENERIC DRUGS
Intense discussion is underway that should clarify the regulatory requirements for generic marketing, at least in Europe. The central issue is at what point generic manufacturers are entitled to make reference to an innovator’s clinical trials data to support a marketing authorisation application for a copy product. This issue will soon come to the fore in acute migraine therapy.

VI. EXAMPLES OF LANDMARK WELL-DESIGNED TRIALS

VII. SUGGESTED READING

Chapter 18. Alzheimer’s Disease and Other Dementias

Ettore Beghi, M.D. ¹

Giancarlo Logroscino, M.D., Ph.D. ²

Amos Korczyn, M.D., M.Sc. ³

¹ Department of Neurology
University of Milano-Bicocca
Monza
ITALY
and
Institute for Pharmacological Research “Mario Negri”
Milano
ITALY

² Department of Epidemiology
HSPH, Harvard University
Boston, MA
USA

³ Department of Neurology
Tel Aviv University Medical School
Tel Aviv
ISRAEL
I. INTRODUCTORY REMARKS

Disease definition, causes, and frequency

The term dementia refers to a syndrome characterized by progressive loss of cognitive functions, including memory and language, and changes in personality and behavior. The most common cause of dementia in the elderly is Alzheimer’ disease (AD) (about 60%), followed by vascular dementia (VaD) (20%). However, considerable overlap exists between AD and VaD, and these two disorders may be extremes of a spectrum. In particular, the clinical differentiation between these disorders is difficult. AD is characterized by the presence of major impairment in learning and in retaining new information and at least one of the following: impairment of complex tasks, impaired reasoning ability, impaired spatial ability and orientation, and impaired language. In VaD loss of cognitive functions roughly overlaps that of AD, except for the onset or worsening of the symptoms within three months of a stroke and/or presence on neuroimaging of brain infarctions involving cortical or subcortical structures, including white matter. Dementia must be distinguished from mild cognitive impairment (MCI), which is a new memory complaint, preferably corroborated by an informant, with objective evidence of impairment of short-term memory, all other cognitive functions being normal, and no substantial interference with daily living activities. MCI is frequently an intermediate stage between normal cognition and dementia and is a risk factor for dementia.

Dementing disorders are followed by an increasing disability resulting in social and occupational decline. Dementia may also occur in younger persons where the etiology includes brain trauma, schizophrenia, multiple sclerosis and AIDS. In underdeveloped countries other infections of the CNS may also be a common cause of cognitive decline.

Over the past two decades, a series of well-conducted epidemiological studies have shown that dementia is a common condition affecting up to 5% of individuals over 65 years in industrialized countries. Most population-based studies indicate a prevalence increasing with age from 1.5/100 between 65 and 69 years to over 30/100 at age 85-89. It is not clear if after age 80-85 the prevalence of the disease is still rising or it tends to stabilize. The incidence of dementia is about 1/100 per year after age 65. In 1997 there were about 2.3 millions individuals with AD in the U.S. (range 1.1 to 4.6 millions) where the number of new cases is about 360,000 each year. Due to the aging of the population, the prevalence of AD in the US will grow about 4-fold within the next 50 years if effective interventions to delay the onset of disease are not developed.

Goals of treatment

There are several goals of the treatment of dementia, including: 1) prevention of disease occurrence; 2) symptomatic improvement; 3) cure of disease; 4) delaying cognitive decline; 5) complete symptomatic control. At present there is no proven treatment that modifies the natural history of the disease or changes its outcome. Potential areas for intervention include: 1) increasing the levels of neurotransmitters involved in cognitive functions; 2) providing neuroprotection to neurons already damaged or showing functional changes; and 3) neuronal regeneration by replacing neurons which have been lost. In the future the goal will be to reverse the natural history of dementia by reversing the typical neuropathological lesions (ß-amyloid and tau).

Symptomatic treatment of dementia

The first paper describing an effective symptomatic treatment of dementia was published less than twenty years ago and was a small cross-over study with tacrine, the first cholinesterase inhibitor (ChEI). In the last two decades, several drugs have been tested and approved. Patients with dementia of any severity treated in randomized clinical trials with donepezil, rivastigmine, and galantamine (which represent the main category of ChEI shown to be effective with an acceptable tolerability profile) experience some benefits in cognitive function, activities of daily living and behavior at least during the first year of treatment. Memantine, an NMDA-receptor antagonist, and possibly ginko biloba have also been reported to be effective.
Problems with trials on symptomatic treatment of dementia

To the present time, the efficacy of symptomatic drugs is at best modest. However, the results of the published randomized trials must be interpreted in the light of methodological drawbacks, concerning especially the definition and the choice of the appropriate outcome. Many outcome measures have been used in dementia trials. The ideal outcome should be easy to measure and easily collectable at each follow-up over a significant period of time (ideally for several years). These measures should have a good reliability, especially considering that many trials are multicenter-based. Cognitive decline is difficult to measure with the available instruments in a quantitative way. Typical problems are floor and ceiling effects, regression to the mean, learning effects and placebo factors. Moreover the range of changes over a short period is small compared to the possible cognitive range of each scale. More suitable and robust end-points are the following: 1) loss of independence; 2) loss of a specific daily living function; 3) placement in a nursing home. However, these end-points require prolonged follow-up and may be influenced by other environmental factors like the presence of an active caregiver, and economic and social environment.

Cognitive tests include a battery of tests covering memory and other domains including language, constructional abilities, attention/concentration and psychomotor speed. Remote and recent memory must be extensively explored along with recall and recognition for various modalities. Verbal and visuo-spatial memory must be also investigated. The Mini Mental State Examination (MMSE) is generally used as a screening test for cognitive impairment while the Clinical Dementia Rating (CDR) is used for grading disease severity. Although the Alzheimer’s Disease Assessment Scale cognitive subscale (ADAS-cog) fulfills the requirements of a comprehensive cognitive scale, none of the available instruments can be preferred as being more valid and reliable. Although several scales have been proposed to measure the activities of daily living (ADL) and to assess an overall clinical improvement, none has advantages that would justify its preferred use in regulatory trials. Moreover, all these tests have shown acceptable validity and reliability in AD, but not in other dementing disorders. This is of particular concern in some aspects. For example, the MMSE is heavily weighted to memory and left hemisphere functions, e.g. speech.

Criteria for assessing symptomatic improvement

The Aging-Warner consortium established in 1992 that two types of outcomes should be used to assess the efficacy of an anti-dementia product: 1) a global assessment performed by a skilled clinician; 2) a performance-based objective test of cognitive function. The consortium identified the Clinician’s Global Impression Scale (CGI) and the Scale ADAS-cog as two preferable instruments. This two-outcome approach has the goal to identify changes that are at the same time clinically meaningful and specific (as determined by cognitive testing) and that are not due to some non-specific effects on the general clinical state. In line with the Note for Guidance on Medicinal Products of the European Agency for the Evaluation of Medicinal Products (EMEA), the efficacy criteria for a drug tested for the treatment of dementia may include symptomatic improvement (which may be manifest in enhanced cognition, more autonomy and/or improvement in behavioral dysfunction), slowing or arrest of symptom progression, and primary prevention of disease by intervention at a pre-symptomatic stage. The EMEA requires a six-month treatment for demonstration of efficacy and one year for maintenance of efficacy. Improvement of symptoms should be assessed in three specific domains: 1) cognition, as measured by neuropsychological tests (cognitive end-point); 2) activities of daily living (functional end-point); and 3) overall clinical response (global end-point). Efficacy variables should be defined for each domain. Primary efficacy variables should include cognitive end-points and the clinical relevance of the improvement in cognition, measured preferably by a functional end-point. The overall benefit should then be measured in terms of the proportion of patients achieving a meaningful benefit (responders). Other end-points include behavioral symptoms, which may be selected as secondary end-points or primary end-points in trials designed to assess control of behavioral abnormalities. The Food and Drug Administration (FDA: http://www.fda.org) has similar guidelines with a two-outcome approach to assess the efficacy of a drug (one cognitive and one global assessment measure). The FDA requires three-month duration with at minimum 1000 patients exposed for several weeks in the relevant dosage range.
Prevention of dementia
The methodological concerns with symptomatic treatment trials (e.g., cognitive function testing) also apply to prevention trials. RCTs for prevention require very large sample sizes to ensure enough power to detect a significant reduction in the incidence of dementia. Because of this problem, most of the data available to date are based on observational studies. A possible strong end-point could be the clinical diagnosis of dementia (incidence of disease) in previously non-demented subjects. A clinical diagnosis of MCI has more problems in terms of definition and reliability. An additional problem of prevention trials is the long duration of follow-up necessary. The selection of subgroups at higher risk can make it easier to reach the numbers needed. Higher risk groups could be subjects carrying APO-E4, subjects in older age groups, or subjects with MCI. However, results from studies in such selected groups may not necessarily be generalized to wider populations. The incidence of AD is influenced by several factors, such as age, sex, education, family history of disease and genetic status (APO-e), which may affect the outcome of a trial. In a recent study, Kriyscio and coworkers (1) revised the factors affecting the probability of AD in a preventive trial: length of the follow-up period, accrual, incidence rate of disease, drop-in, drop-out rate and the adherence rate. PREADVISE, which is the largest prevention trial up to date, focuses on assessing the potential protective effect of selenium and vitamin E and plans to enroll 10,700 subjects at 400 sites. Another factor that can increase sample size is the rate of misdiagnosis, especially for mild new cases that could not be identified with the operational diagnostic instruments of the trial.

The use of placebo in dementia trials
Given the great heterogeneity of dementing disorders in terms of symptom profile, overall severity and course, the efficacy of a drug can be demonstrated only by using appropriate controls. The availability of active treatments raises ethical concerns about the use of placebo in patients with dementia. However, use of placebo may still be considered for several reasons: 1) Drug efficacy could possibly be only detected with placebo as a comparator; because superiority against an active control may be difficult to prove. Equivalence or non-inferiority designs (requiring larger samples) may be used as proof of efficacy in studies using active controls, but such studies raise interpretative problems, because the argument could be raised that the reference control was not necessarily efficacious in the population and under the conditions being studied; 2) There is no evidence that available treatments affect the patient’s long-term health; 3) The efficacy of cholinesterase inhibitors is at best modest; 4) Lack of effective drugs for the prevention of dementia justifies the use of placebo to test drugs assessed for this indication; 5) There are difficulties in determining rates of adverse effects against active comparators which may also cause adverse effects.

The Quality Standard Subcommittee of the American Academy of Neurology (2) recognized that the use of cholinesterase inhibitors should be considered a standard of care for AD patients, even if the clinical effect is small. This requirement may justify placebo-controlled trials only when the investigational agent is used as add-on to existing treatments.

Issues with informed consent
Putative anti-dementia drugs typically undergo a staggered development process, culminating in double-blind studies, usually with a parallel design in which placebos are employed. Patients asked to participate in controlled trials of an investigational treatment must be informed of any alternative treatment and should be able to explore the positive and negative consequences of the treatments being tested (including the use of placebo) to provide a fully informed consent.

Participation of demented patients in such studies calls for attention to the special circumstances of this population and their needs, including consideration of potential benefits to the patients as well as caregivers, the economic impact, and the expected benefits to the society and science. However, demented patients may not be able to understand the full implications of the study, and may be unduly influenced by researchers and caregivers. In addition, the practice of obtaining proxy consent from the patient’s surrogate does not satisfactorily resolve the ethical issues, as the surrogate’s decisions usually reflect their
personal rather than the patient’s choice. In summary, the inability of patients to fully comprehend the possible implications of the study and the consequent need for the consent to be provided by a surrogate raises notable problems. The investigators and the IRB have an important role in ensuring that drug studies for patients with dementia are performed in a way that provides optimal information and preserves the well-being of patients as well as support for their caregivers.

**Biological markers**

Biomarkers can be defined as biological compounds that can be measured as indicators of exposure (risk factors of disease), intermediate steps of pathogenic pathways, or different clinical stages of the disease including specific responses to drug therapy. Genetic markers (e.g. the apolipoprotein E genotype) can help to identify subjects who are at higher risk of dementia. Several blood and CSF tests have been proposed for the early detection of AD. Such markers should reflect the pathophysiological mechanisms of AD, an example being the measurement of brain, serum or CSF β-amyloid and tau protein concentrations to detect altered metabolism of amyloid and neurofibrillary degeneration. Tau is a microtubule-associated protein that forms the basic element of the neurofibrillary tangle, one of the characteristic lesions of AD. CSF-tau levels have shown a good sensitivity (85%) and specificity (83%) to distinguish AD patients from normal elderly controls. Aβ42 is a 42 aminoacid fragment of the transmembrane amyloid precursor protein (APP) that aggregates as β-pleated sheets in extracellular neuritic plaques. Aβ42 and a shorter 40-amino-acid peptide (Aβ40) can both be assayed in the CSF. Several studies have consistently demonstrated a moderate to marked decrease in CSF Aβ42 in AD, probably because this compound is bound within the neuritic plaques.

At present, no biological marker has been recommended for use in clinical trials of dementia. The use of biological markers could be useful in the future, especially among subjects who are asymptomatic or have MCI, to select subgroups at higher risk to develop dementia. Alternatively, biological markers can be considered as surrogate end-points to assess treatment efficacy. A consensus committee has recently proposed a reclassification of biomarkers for AD in clinical practice (3). These include core markers (those judged to have reasonable evidence for association with key mechanisms of AD pathology) and non-core markers (those felt to be less clearly associated with mechanisms of pathogenesis or neurodegeneration in AD). Core markers include amyloid beta peptide, APP, tau proteins, isoprostanes, A1-antichymotrypsin, interleukin-6-receptor-complex, C-reactive protein, C1q, homocysteine, oxysteroids, 3-nitrotyrosine. Non-core markers include glutamine synthetase, human antibodies against Aβ-related proteins, glial fibrillary acidic protein, sulfatide, AD7C/NTP, and kallikrein 6.

**II. PHASE II STUDIES FOR REGISTRATION OF NEW SYMPTOMATIC DRUGS**

**II.1. Outline of a typical development plan**

During this phase the candidate drug is tested against placebo. The goal of this phase is to document efficacy and to identify the parameters of the treatment regimen (titration, dose regimen, maximal tolerated dose, etc.) most likely to maximize the therapeutic response in patients with well-defined disease. As a rule, phase II studies are designed to maximize efficacy by using the smallest possible number of patients with homogeneous disease characteristics, notably those with fewer concomitant illnesses and less severe impairment, in whom clinical response can be detected over a relatively short time. To increase sensitivity, drug response may also be tested by enrolling patients who responded during a pre-randomization phase. Phase II controlled trials must be preceded by open label exploratory studies to assess titration rates, maximally tolerated doses and pharmacokinetics.

The typical trial design is randomized, placebo-controlled, parallel group testing at least two dose regimens over a short time period. Titration to the predetermined doses should be identified to minimize...
drop-outs for adverse effects. Patients included in short term phase II clinical trials should be allowed to participate in long-term trials.

II.2. Short-term phase II studies

II.2.A. Objectives
To evaluate short-term efficacy and tolerability and to detect a correlation between different doses and positive and untoward effects.

II.2.B. Primary end-points
1. Change in cognitive function (measured by psychometric tests)
2. Clinical global impression of change
3. Change in performance of ADL
4. Acceptability of treatment as measured by withdrawal from trial
5. Safety as measured by incidence of adverse events, particularly those leading to withdrawal

II.2.C. Secondary end-points
1. Behavioral disturbances
2. Change in quality of life
3. Effect on caregiver

II.2.D. Exploratory end-points
1. Plasma drug levels
2. Changes in functional imaging
3. Effects on biological markers

II.2.E. Study design
Multicenter, randomized, placebo-controlled, parallel group. A cross-over design may be employed in short-lasting treatment periods and when carry-over effects are insignificant. A screening phase may be used to verify patient eligibility, followed by a prospective baseline period during which cognitive functions are tested and the functional and global clinical activities are measured. After randomization, a titration period is started of sufficient length to achieve steady state conditions. A maintenance period of six months follows under the assumption that during this period there will be a clinically significant progression of the disease (for example, a 4-point change on the ADAS-cog) in the placebo arm. Eligible patients should be free of concomitant illnesses and taking no or few active principles. At the end of the trial, the patient is either withdrawn according to a pre-defined treatment schedule, or he/she enters a long-term phase.

II.2.F. Planned sample
With two treatment arms (active vs placebo), a sample size of about 100 per treatment group is needed with an expected 15% of responders (e.g. a 4-point or greater improvement on the ADAS-cog) in the placebo group, under the assumption to detect a 20% absolute difference in the proportion of responders, with an 80% power, a 5% (two-sided) significance, and a 20% drop-out rate.

II.2.G. Study population
Patients with definite dementia, AD or VaD.

II.2.H. Specific inclusion criteria
a. Adult female and male patients
b. MMSE between 10 and 26
c. Reliable caregiver
d. Dementia of mild to moderate severity (CDR<3 within 4 weeks prior to entry).
e. Imaging studies performed during six months prior to entry consistent with the diagnosis.

II.2.1. Specific exclusion criteria
a. Delirium or impairment of consciousness
b. Major depression or other significant psychiatric diagnosis
c. History of drug or alcohol abuse
d. Other disorders possibly causing dementia
e. History of hypersensitivity to relevant drugs.
f. Neoplastic, hepatic, renal or cardiac disorders of significant impact on function or survival
g. Any disability preventing compliance with test procedures.

II.2.J. Tools for assessing primary end-points
a. ADAS-cog or other valid and reliable cognitive scale;
b. CGIC, CIBIC plus or other valid and reliable scale assessing overall clinical impairment

II.2.K. Tools for assessing secondary end-points
a. PDS, IADL or other valid and reliable scale assessing ADL, quality of life, cognitive and behavioral abnormalities
b. Caregiver global impression
c. Nurse global impression

II.2.L. Specific criteria for early withdrawal and discontinuation
a. Occurrence of significant adverse events thought to impair ADL and overall quality of life
b. Poor compliance
c. Withdrawal of consent

II.2.M. Data analysis method
The analysis of treatment efficacy is performed on the intention-to-treat population (all randomized patients receiving at least one dose of study medication). Parametric and non parametric tests are used as appropriate for primary and secondary end-points. Continuous variables (cognitive scores) are tested using parametric tests, like the Student’s t test or analysis of variance, and non parametric tests, like the Wilcoxon-Mann Whitney test. Categoric variables (global impression, ADL, IADL, proportion of responders or cases withdrawn from the study, etc.) are tested using the chi-square test (parametric) or the Kruskal-Wallis test (non parametric). Changes in the rate of decline can be assessed with survival analysis (Kaplan-Meier survival curves and Cox’s proportional hazard function). Univariate and multivariate statistical techniques can be used as appropriate. All p values should be based on two-sided tests with a 5% significance level.

II.3. Long-term phase II studies
When treatment efficacy is demonstrated in short-term clinical trials, long-term phase II studies are implemented to verify the duration of treatment effects on cognitive, functional and behavioral parameters. The estimated duration of a long-term clinical trial is about 12 months. During this period safety and efficacy are investigated with respect to symptom relief, slowing of progression of cognitive decline, and control of behavioral abnormalities. Ideally, long-term clinical trials are the extension of short-term studies. The dosage of the study medication may be adjusted to achieve the maximally tolerated dose. At study end all patients, including those who were in the placebo arm, should be given the active medication, which should be continued as long as the physician and/or caregiver perceives it to be beneficial, and retention time should be used as a measure of treatment effectiveness and tolerability. Along with cognitive tests, treatment benefits should be measured in terms of effects on hard end-points like time to loss of independence and/or relevant functional impairment.
II.3.A. Objectives
To evaluate long-term efficacy and tolerability of treatment

II.3.B. Primary end-points
a. Change in cognitive function (measured by psychometric tests)
b. Clinical global impression of change
c. Change in performance of ADL
d. Time to loss of independence and/or relevant functional impairment
e. Retention time as a measure of treatment efficacy
f. Acceptability of treatment as measured by withdrawal from trial
g. Safety as measured by incidence of adverse events leading to withdrawal

II.3.C. Secondary end-points
a. Behavioral disturbances
b. Change in quality of life
c. Effect on caregiver

II.3.D. Exploratory end-points
Effects on biological markers

II.3.E. Study design
Multicenter, randomized, placebo-controlled, parallel group.

II.3.F. Planned sample
Sample size should be calculated on end-points like loss of independence (placement in nursing home) or severe functional impairment; given a 20% expected 12-month rate in the control group, a sample of about 400 per treatment group is needed under the assumption to detect a 10% difference in the proportion of patients achieving the end-point, with a 80% power, a 5% (two-sided) significance, and a 25% drop-out rate.

II.3.G. Study population
As indicated for short-term phase II studies

II.3.H. Specific inclusion criteria
As indicated for short-term phase II studies

II.3.I. Specific exclusion criteria
As indicated for short-term phase II studies

II.3.J. Tools for assessing primary end-points
As indicated for short-term phase II studies

II.3.K. Tools for assessing secondary end-points
As indicated for short-term phase II studies

II.3.L. Specific criteria for early withdrawal and discontinuation
As indicated for short-term phase II studies

II.3.M. Data analysis method
As indicated for short-term phase II studies
III. PHASE III STUDIES FOR REGISTRATION OF NEW SYMPTOMATIC DRUGS

III.1. Outline of a typical development plan

During phase III development at least one large multicenter, randomized, double-blind, placebo-controlled, parallel-group confirmatory trial must be undertaken. A factorial design can also be considered comparing the investigational drug with cholinesterase inhibitors and placebo. Patients to be enrolled should have a definite diagnosis, like probable AD or VaD. In contrast to phase II trials, an effort should be made to include patients representative of those usually seen in clinical practice. As well, the choice of daily drug doses and dose increments should follow the patterns of clinical practice. As with phase II studies, outcome measures must include changes in cognitive scores and functional and behavioral changes. As the goal of phase III studies is to provide evidence of sustained treatment efficacy over a prolonged time period, the estimated length of the study should be about 12 months during which a statistically significant and clinically relevant difference should be documented in favor of the experimental treatment. In studies employing a factorial design, an additive effect is also searched in favor of patients receiving the investigational drug and the other active treatment. Having tested the daily dosage with the best therapeutic ratio in phase II trials, in these pivotal studies the drug should be titrated upwards to reach the highest tolerated dose. At the end of the double-blind phase, an open label extension period should also be considered to test the drug over an even longer period of time and under conditions most likely to reproduce the setting of clinical practice.

III.2. Typical phase III study

III.2.A. Objectives
To evaluate sustained efficacy and tolerability in a sample population representative of that seen in clinical practice

III.2.B. Primary end-points
a. Change in cognitive function (measured by psychometric tests)
b. Clinical global impression of change
c. Change in performance of ADL
d. Time to loss of independence and/or relevant functional impairment
e. Retention time as a measure of treatment efficacy
f. Acceptability of treatment as measured by withdrawal from trial
g. Safety as measured by incidence of adverse events leading to withdrawal

III.2.C. Secondary end-points
a. Behavioral disturbance
b. Change in quality of life
c. Effect on caregiver

III.2.D. Exploratory end-points
Effects on biological markers

III.2.E. Study design
Multicenter, double-blind, placebo-controlled, parallel group with or without factorial design

III.2.F. Planned sample
As with long-term phase II studies.

III.2.G. Study population
Patients with dementia, including probable AD and VaD.
III.2.H. Specific inclusion criteria
   a. Adult female and male patients
   b. MMSE between 10 and 26
   c. Reliable caregiver
   d. Dementia of any severity supported by imaging studies performed during six months prior to entry

III.2.1. Specific exclusion criteria
As indicated for short-term phase II studies

III.2.J. Tools for assessing primary end-points
As indicated for short-term phase II studies

III.2.K. Tools for assessing secondary end-points
As indicated for short-term phase II studies

III.2.L. Specific criteria for early withdrawal and discontinuation
As indicated for short-term phase II studies

III.2.M. Data analysis method
As indicated for short-term phase II studies.

IV. OTHER STUDIES

IV.1. PREVENTION TRIALS IN MILD COGNITIVE IMPAIRMENT (MCI)

IV.1.A. Outline of a developmental plan
As patients with MCI are expected to convert to dementia at a rate of about 10-15% per year, they represent the ideal target for a prevention trial. Using as an end-point the conversion to dementia diagnosed according to the DSM-IV or NINCDS-ADRDA criteria, a three-year trial has sufficient length to document a statistically significant, clinically relevant, and sustained treatment effect. Placebo must be used to detect the effects on disease progression attributable to active treatment. This procedure is however not without problems. MCI patients who convert to AD within 3 years are very likely to have the disease already at baseline, and thus the study is not really designed for prevention of dementia but rather examines the effect of the drug on the rate of cognitive decline. A practical issue is that MCI patients who are recruited do not necessarily develop dementia at the high rate reported in the literature, and may not represent the real world of MCI. In particular, these subjects may have a more benign course because they have a more prolonged course or have higher prevalence of anxiety or depression.

IV.1.B. Representative trial protocol

IV.1.C.i. Objectives
To assess the treatment effects on disease progression and conversion to dementia

IV.1.C.ii. Primary end-point
Time to the diagnosis of dementia.

IV.1.C.iii. Secondary end-points
Time to the diagnosis of AD, VaD, and other dementia types.
IV.1.C.iv. Exploratory end-points
Effects on biological markers.

IV.1.C.v. Study design
Multicenter, double-blind, placebo-controlled, parallel group with or without factorial design.

IV.1.C.vi. Planned sample
Given the expected 15% 12-month conversion rate on which the treatment response is measured, a sample of about 120 subjects per treatment group is needed under the assumption to detect a 20% in the proportion of responders, with an 80% power, a 5% (two-sided) significance, and a 30% drop-out rate.

IV.1.C.vii. Study population
Patients with MCI, i.e. individuals with a new deficit in at least one cognitive domain (usually recent memory) but who appear to function independently in ADL.

IV.1.C.viii. Specific inclusion criteria
a. Adult female and male patients
b. Presence of a memory complaint, preferably corroborated by an informant
c. Objective evidence of impairment of short-term memory (for age)
d. Otherwise normal cognitive functions
e. No interference with work, social activities, or other ADL
f. MMSE > 26 (Absence of dementia)

IV.1.C.ix. Specific exclusion criteria
a. Major depression, anxiety or other significant psychiatric diagnosis
b. History of drug or alcohol abuse
c. Other disorders possibly causing cognitive decline
d. History of hypersensitivity to relevant drugs
e. History of (active) neoplasm, hepatic, renal or cardiac disorders, which could affect the patient’s survival
f. Any disability preventing compliance with test procedures

IV.1.C.x. Tools for assessing primary end-points
a. NINCDS-ADRDA criteria for the diagnosis of probable dementia

IV.1.C.xi. Tools for assessing secondary end-points
b. DSM-IV criteria for the diagnosis of dementia
c. NINDS-AIREN criteria for the diagnosis of vascular dementia, Hachinski Ischemic Scale for the assessment of vascular dementia
d. Work Group on Frontotemporal Dementia and Pick’s Disease diagnostic criteria
e. Consensus Guidelines for the diagnosis of dementia with Lewy bodies

IV.1.C.xii. Specific criteria for early withdrawal and discontinuation
a. Occurrence of serious adverse events or events thought to impair ADL and quality of life
b. Poor compliance
c. Withdrawal of consent.

IV.1.C.xiii. Data analysis method
The analysis of treatment efficacy is performed on the intention-to-treat population. Univariate and multivariate statistical techniques can be used as appropriate (see also short-term phase II studies).
Conversion to dementia (in general and by type) can be assessed with survival analysis (Kaplan-Meier survival curves and Cox’s proportional hazard function). All p values are two-sided with a 5% significance level.

**IV.2. PREVENTION TRIALS IN ASYMPTOMATIC PATIENTS**

**IV.2.1. Outline of a developmental plan**

A trial on the prevention of dementia in asymptomatic elderly individuals must be designed considering the expected incidence of dementia in the study population and the factors most likely to affect the incidence of the disease. These factors include age, family history of disease, race/ethnicity, education, and genetic background. In addition, the expected number of patients developing dementia depends on the length of the follow-up period and the drop-out rate (mostly caused by death and poor compliance). Several assumptions are thus required for the calculation of the accrual period, the estimate of the sample size, and the duration of the follow-up.

**IV.2.2. Representative trial protocol**

**IV.2.A.i. Objectives**

To assess the treatment effects on the incidence of dementia.

**IV.2.A.ii. Primary end-point**

Reduction of the incidence of dementia.

**IV.2.A.iii. Secondary end-points**

Reduction in the incidence of dementia in patient subgroups defined by age, education, and genetic factors.

**IV.2.A.iv. Exploratory end-points**

Effects on biological markers.

**IV.2.A.v. Study design**

Multicenter, double-blind, placebo-controlled, parallel group with or without factorial design.

**IV.2.A.vi. Planned sample**

Based on the calculations made for the PREADVISE study, a sample of about 2700 individuals per treatment group is required to halve the incidence of dementia with a 80% power, a 5% level of significance, and a 30% drop-out rate.

**IV.2.A.vii. Study population**

Asymptomatic elderly individuals.

**IV.2.A.viii. Specific inclusion criteria**

- Female and male patients aged 65 years and older
- Normal cognitive functions

**IV.2.A.ix. Specific exclusion criteria**

- Major depression, anxiety or other significant psychiatric diagnosis
- History of drug or alcohol abuse
- Other disorder possibly causing cognitive decline
- History of hypersensitivity to relevant drugs
e. History of (active) neoplasm, hepatic, renal or cardiac disorders, which could affect the patient’s survival
f. Any disability preventing compliance with test procedures

IV.2.A.x. Tools for assessing primary end-points
NINCDS-ADRDA criteria for the diagnosis of probable dementia

IV.2.A.xi. Tools for assessing secondary end-points
Genetic background can be defined by history taking and APO-E genotype

IV.2.A.xii. Specific criteria for early withdrawal and discontinuation
a. Occurrence of serious adverse events or events thought to impair ADL and quality of life
b. Poor compliance
c. Withdrawal of consent.

IV.2.A.xiii. Data analysis method
The analysis of treatment efficacy is performed on the intention-to-treat population. Univariate and multivariate statistical techniques can be used as appropriate (see also short-term phase II studies). Incidence of dementia (in general and by type) can be assessed with survival analysis (Kaplan-Meier survival curves and Cox’s proportional hazard function). All p values are two-sided with a 5% significance level.

IV.3. PRAGMATIC TRIALS
As in other clinical conditions, pragmatic trials are designed to reproduce settings reflecting more closely the use of a drug in clinical practice. Pragmatic trials may be designed to assess the effectiveness of different therapeutic strategies (e.g., early vs delayed treatment) and to test treatment in populations usually not included in regulatory trials (e.g., oldest patients or patients with concurrent disabling disorders). Studies comparing different drugs and allowing dosing flexibility could be considered. Survival analysis with retention time as the primary end-point should be the preferred choice for measuring treatment effectiveness. Other outcome measures could include time to nursing home placement or loss of independence. An intent-to-treat analysis should be performed in all cases. Trial duration may vary according to the type of therapeutic strategy but it should be generally no shorter than 24 months.

IV.4. SPECIAL INDICATIONS
The designs described for phase III clinical trials can also be used for the assessment of efficacy in patients with specific syndromes (dementia associated with cerebrovascular disorders, dementia with Lewy bodies, fronto-temporal dementia). In these cases, the inclusion/exclusion criteria, the primary and secondary end-points, and the relative tools are the same as those used for dementia at large. As with the management of concurrent clinical conditions, the treatment of the underlying disorders (stroke, Parkinsonism, etc.) should be carefully considered in terms of interactions and specific contraindications.

IV.4.i. Studies in Vascular Dementia

Vascular dementia (VaD is considered the second most common form of dementia after AD worldwide but probably the first in some countries. Cerebrovascular disease can determine VaD with different mechanisms including large-vessel disease with multiple strokes, single strokes in strategic areas, or subcortical lesions with multiple lacunar infarcts and white matter lesions. The diagnosis of VaD is possible when dementia, history of cerebrovascular disease and a relationship between the two disorders is present. The characteristic feature of subcortical VaD is the involvement of executive functions. These include ability to execute complex behaviors and solving-problems ability. MMSE is not a good
instrument to assess executive functions. Several tests (among which the Trail-Making test or the Clock Drawing task) can assess executive functions. These tests should be included in the assessment of the diagnosis and follow-up of VaD.

Several risk factors are associated with cerebrovascular diseases and consequently with VaD.

The prevention of cerebrovascular disease should be the first step in the prevention of VaD. Studies looking at the efficacy of controlling cardiovascular risk factors in the prevention of VaD are few. In the non-demented subjects enrolled in the Syst-Eur study who received antihypertensive treatment the risk of dementia was less than 50% of that of controls (4).

There is growing evidence that in VaD as in AD there is involvement of the cholinergic system. Animal models of stroke-prone spontaneously hypertensive rats present behavior that can be considered similar to memory impairment present in VaD. These rats show reduction of acetylcholine in several areas of the brain including hippocampus. Cholinergic agents have therefore been tested as potential treatments in VaD. In the largest trial, 603 subjects were recruited for a multicenter randomized trial on donepezil in VaD (5). Patients with probable and possible VaD were recruited. Donepezil was found to be effective for patients with VaD using the ADAS-cog and CIBIC –plus as outcome measures.

IV.4.ii. Behavior and mood disturbances in demented subjects

Depression, anxiety, agitation and more serious symptoms like delusions and aggressive behavior are commonly seen in AD. Dementia-related behavioral disturbances have been associated with excess disability, increased caregiver burden, and premature institutionalization. The presence of behavioral disturbances is one of the main reasons for exclusion of patients from clinical trials. Pharmacotherapy is often necessary to treat these disturbances and specific trials are indicated to document efficacy and tolerability for this specific indication. For example, a NIH-sponsored trial is currently recruiting AD patients to study the effects of citalopram and risperidone in people with dementia-related behavior problems.

V. EXAMPLES OF LANDMARK WELL-DESIGNED TRIALS

postmenopausal women: the Women's Health Initiative Memory Study: a randomized controlled trial. JAMA 2003;289:2651-2662.

VI. SUGGESTED READINGS


VII. REFERENCES

Chapter 19. Parkinson’s Disease and Other Extrapyramidal Disorders

Caroline M. Tanner M.D., Ph.D. 1

Christopher G. Goetz, M.D. 2

Yoav Ben-Shlomo, MBBS, Ph.D. 3

Piu Chan, M.D., Ph.D. 4

1 Parkinson’s Institute
1170 Morse Avenue
Sunnyvale, CA 94089
USA
Tel: 408-734-2800
Fax: 408-734-8455

2 Rush Medical Center
Chicago, IL
USA

3 Department of Social Medicine
University of Bristol
Bristol
UK

4 Xuan Wu Hospital of Capital University of Medical Sciences
Beijing
CHINA

Corresponding Author:
Caroline M. Tanner M.D., Ph.D.
I. INTRODUCTORY REMARKS

Parkinson’s disease (PD) is the second most common neurodegenerative disorder of later life. The primary areas of neuropathologic injury are pigmented aminergic neuronal populations - the substantia nigra, locus ceruleus and others. PD causes progressive disability. Disability associated with PD includes not only motor dysfunction, but for many dysautonomia, cognitive changes (ranging from loss of executive function to frank dementia) and depression. There is no known cure nor is there a recognized method for slowing the ongoing degenerative process.

I.1. Unfulfilled therapeutic needs

Currently available therapies are most effective in minimizing the motor dysfunction of PD, through increasing the activity of the nigrostriatal dopamine system, although overall benefit is partial, and not sustained over the years of disease. Development of a treatment with sustained therapeutic benefit over the decades of the disease would be invaluable. In addition, these standard therapies also have disabling and/or dose limiting adverse effects. Acute adverse effects include nausea and hypotension. Chronic adverse effects include dyskinesias (involuntary movements), somnolence, hallucinations or psychosis. In more advanced disease, a progressive diminution in the overall motor benefit, and in the duration of the treatment effect is an additional limitation of some therapeutic approaches. Methods for preventing the development of these disabling adverse effects, or for treating them effectively once developed, are needed. Treatments for the nonmotoric disabilities of PD (such as dysautonomia, cognitive changes and depression) are few, and none are universally effective. Developing effective therapies for these nonmotor features of PD is an emerging area of interest.

Much effort has been directed to the development of treatments that can stop or slow the progression of established disease (neuroprotection), but so far such an agent has not been identified unequivocally. Similarly, no cure is available, and there is no way to delay the onset of PD. These latter questions are the most critical, since they address preventive or curative goals, rather than symptom amelioration.

Other related areas of importance involve pharmacogenetic investigations in PD and developing treatments for “atypical” parkinsonism. Little work has been done to determine whether the genetic makeup of the individual in part determines response to therapy. This would include investigating subgroups with genetic forms of parkinsonism, as well as PD patients with genetic variants of elements of the dopaminergic system (such as metabolic enzymes or receptors). The “atypical” parkinsonian syndromes include less common chronic neurodegenerative disorders with prominent parkinsonian features, such as multiple system atrophy and progressive supranuclear palsy. In contrast to PD, there are no effective therapies for these devastating disorders, although existing antiparkinsonian therapies may provide short-lived partial benefit for a minority of patients.

I.2. Unclarified issues related to current treatments

The majority of studies of antiparkinsonian agents have compared single agents to placebo in order to demonstrate efficacy. Few controlled studies provide evidence to guide the choice of a treatment regimen among the many existing therapies. Controlled studies have rarely compared existing therapies of any type, either those with proposed neuroprotective effects or those with symptomatic effects, either alone or in combination. In addition, little is known regarding the effects of existing therapies within clinical subgroups, defined by demographic characteristics such as age, gender or race/ethnicity or by disease features such as tremor predominance, cognitive function or age at onset. Little is known regarding the benefit of any therapeutic agent for a period of more than a few years. A review of all trials up until the end of 2001 found that the median follow-up period per trial was two years and only 40% of trials in early PD went beyond 12 months. Only two trials (DATATOP and UKPDRG) followed up patients for up to 10
years with only the latter being designed to test differences in mortality. Whether the choice of therapeutic regimen can prevent the development of adverse effects or alter the course of disease remains unresolved. Almost nothing is known regarding the relative effects of any individual therapy on survival in PD.

I.3. Needs and justification for developing new drugs

Despite the number of agents approved for use in PD, neither a risk-free treatment with sustained benefit nor a preventive or curative agent has been identified. Therefore, there is a need for new drug development in all aspects of PD therapy. Moreover, because PD and other late-life neurodegenerative disorders, such as Alzheimer’s disease, are thought to share common pathogenic features, there may be potential overlap of effect in some areas – either for primary neuroprotection, or treatment of common symptoms such as cognitive impairment or depression. Moreover, the potential number of cases of PD worldwide is expected to increase with the aging of the population. Most studies from the USA and Western Europe indicate PD prevalence to range from 1 – 2 % among individuals 65 and older, and age-specific prevalence appears to double about every 5-7 years after the age of 65. The numbers of persons in the age group at risk for PD is expected to increase progressively over the next several decades in both the developed and developing world. This expected increase in the numbers of persons affected is also expected for other neurodegenerative disorders such as the atypical parkinsonisms and Alzheimer’s disease. A second, unrelated area of potential overlap for dopaminergic agents is in the treatment of restless legs syndrome, estimated to affect between 5 – 15% of the adult population.

I.4. Particularities of PD that will influence the protocol of investigation of the drug

There is no diagnostic test for PD, and diagnosis is based solely on the expertise of the examiner. The greater the expertise of the examiner, the more accurate the diagnosis, as compared to post-mortem examination, but even in the hands of experts, some diagnostic error is expected. This potential for error is greater when the PD is of short duration, and clinical signs are few. Because PD is disabling, most persons can function without therapy only within the first one or two years after diagnosis. This hampers the evaluation of agents proposed to slow, but not stop, disease progression, as the addition of symptomatic therapy confounds the evaluation of a neuroprotective effect judged by clinical criteria. Recently, imaging approaches targeting the nigrostriatal dopamine system have been proposed as adjunctive means for assessing progression of PD, although these, too, may not be independent of the potentially confounding pharmacologic effects of PD treatments. A second aspect of the requirement for symptomatic therapy in moderate or more severe PD is that the efficacy of any agent in this patient population must be evaluated as an adjunct to an established therapy, because comparison to placebo alone would not be ethical. However, comparison of a new drug to placebo is appropriate for patients receiving symptomatic therapy for the first time (sometimes called “de novo” patients), and this design is preferred by both the FDA and the EMEA for new drugs intended for registration as monotherapy. In this special setting, provision for “rescue” with a symptomatic agent may be advisable.

The response to some antiparkinsonian agents, notably those including l-dopa, can vary dramatically over time. Variation can occur over hours or even minutes during the course of a single dose, and can be further modified by conditions such as the time of day, the number of prior doses, the type of concurrent therapy, and the timing and protein content of meals. Similarly, the adverse effects of antiparkinsonian therapies, such as dyskinesias, typically wax and wane during the course of a day. These features must be taken into consideration when choosing the most appropriate outcome measure.

Newer therapies may show modest improvements in motor functions. However, without the collection of data on activities of daily living and quality of life, it is hard to evaluate whether such benefits make much difference to patients and are cost-effective.
II. PHASE II STUDIES FOR REGISTRATION OF NEW DRUGS

II.1. Outline a typical development plan

During Phase II of drug development, candidate therapies are usually tested against placebo, though occasionally other existing treatments, in the disease population of question. Generally, PD clinical trials address two primary impairments, either parkinsonism itself or motor complications that occur as part of the disease and its chronic treatment and take the form of dyskinesias (involuntary movements) or motor fluctuations (poor response to medications). For each, the primary focus can be on the delay of development of impairment or treatment of impairment once developed. Patient selection depends on the primary clinical problem being addressed. For example, studies of delay in clinical progression of parkinsonism focus on early disease and patients are usually on no other medications for Parkinson’s disease, whereas treatment protocols for dyskinesias and motor fluctuation usually focus on patients with more advanced disease who are already on multiple antiparkinsonism drugs. Almost all randomized, double-blind, controlled trials in PD are parallel in design.

II.2. Short-term studies

Short-term efficacy trials are usually three to six months with built-in titration and withdrawal phases. In some of these studies, long-term open label continuation phases are included for the acquisition of safety data.

II.2.A. Objectives

Studies usually are aimed at treating impairment and therefore focus on alleviating parkinsonism itself or improving existing dyskinesias or motor fluctuations.

II.2.B. Primary Endpoints

For parkinsonism:

a. Comparison of the Unified Parkinson’s Disease Rating Scale (UPDRS) total score in relation to baseline scores.

b. Comparison of the UPDRS Motor Examination (Part III) can be used as well, or the combined Activities of Daily Living and Motor Examination score (Parts II and III) in relation to baseline scores.

c. The UPDRS is internationally utilized and has largely replaced earlier scales like the Columbia and Webster scales.

d. The Hoehn and Yahr scale was formerly used, but it is a non-continuous scale, poorly responsive to interventions, and therefore more frequently used currently to describe patient groups and define entry criteria, rather than serving as a primary end-point.

For motor fluctuations:

a. Dyskinesias are usually rated with the Abnormal Involuntary Movement Scale, or the Rush Dyskinesia Scale.

b. Motor fluctuations are measured with at-home diaries for which patients undergo training in the study center on the operational definitions of “ON” (good medication response), “ON with disabling dyskinesias” (good medication response, but with superimposed involuntary movements that interfere with activities), and “OFF” (poor medication response). Reductions in overall OFF time without an increase in ON with dyskinesias indicate improved motor fluctuations. Global measures on motor fluctuations and dyskinesias can be obtained from UPDRS Part IV.

II.2.C. Secondary endpoints

For parkinsonism:

a. Hoehn and Yahr stage, the Schwab and England rating scale
b. Dyskinesias and motor fluctuation as secondary outcomes are measured as described in Primary endpoints

c. Global scales like the Clinical Global Impression Severity and Clinical Global Impression Change scores are also used. More studies now include disease specific Quality of Life measures, such as the PDQ-39, and PDQUALIF or generic measures such as EuroQol and the SF-36. (It can be argued that quality of life measures should be primary rather than secondary outcomes)

For motor fluctuations or dyskinesias:

a. Secondary endpoints include UPDRS and all primary endpoints listed under Parkinsonism.

II.2.D. Exploratory endpoints
Depending on the chemistry of the intervention, exploratory endpoints can include non-motor elements of Parkinson’s disease such as depression, cognition, hallucinations, and dysautonomia. Rating scales for all of these are derived from standard scales in the medical literature and have not been developed specifically for Parkinson’s disease in most instances.

II.2.E. Study Design
Randomized double-blind placebo controlled parallel studies are the gold standard. After screening and entry criteria are verified, subjects are randomized and are seen regularly during study-drug intervention and then withdrawn from the drug and seen at a close-out visit. Some studies assign patients to a fixed dose (or doses) of the study drug or placebo and others allow dose titration to a maximal tolerated dose that is pre-determined. Patients are usually seen weekly for the first four weeks or through the titration phase, then at longer intervals during chronic treatment with an end-of-exposure visit and a final visit one week after drug-exposure cessation.

II.2.F. Study population
The study population depends on the question being addressed:

Parkinsonism: In early disease, the subjects are usually on no treatment for parkinsonian signs at the time of enrollment and receive monotherapy with either placebo or study drug. In moderate and advanced disease, patients are enrolled who have inadequate efficacy from their current drug therapy and can often be on levodopa or another agent with a pharmacological mechanism that is different from the study drug under question.

Motor complications: In advanced disease where the focus is usually on motor complications, patients must have dyskinesias and/or motor fluctuations of sufficient severity to warrant intervention.

Patients with another illness sufficiently severe as to jeopardize participation in the study or interpretation of results are excluded. Patients must pass screening tests for depression or dementia.

II.2.G. Specific Inclusion Criteria
Parkinson’s disease is defined clinically based on various diagnostic standards, such as the UK Brain Bank criteria. Inclusion criteria for admitting mild, moderate or advanced patients with PD are primarily based on Hoehn and Yahr stage and medication exposure. Early monotherapy studies restrict subjects to Stage I-III (unilateral or bilateral disease that may include mild balance problems but no spontaneous falling) and sometimes to Stage I-II only (no balance problems). Studies of drugs that are added to current treatment in moderate Parkinson’s disease usually restrict patients to Stage II-IV, and therefore include patients with poor balance. For studies of motor complications, inclusion criteria usually require baseline scores for the target problem sufficiently severe enough so that patients are likely to deteriorate during the trial. For dyskinesias, a minimal baseline score on the AIMS (variably 7-10) is often used, and for motor fluctuations, inclusion often requires a minimal 25% or more OFF time on diaries or the UPDRS Part IV for study entry. These criteria are introduced to avoid “floor effects”.

232
II.2.H. Specific Exclusion Criteria
In each trial, patients cannot have an allergy to the product being tested. Current exposure to various medications and past exposure to levodopa may exclude subjects from the early monotherapy trials. Because dopamine is a precursor to melanin, studies of drugs that alter levodopa bioavailability or metabolism exclude patients with a past history of melanoma. Almost no intervention studies permit Stage V patients who are wheelchair-bound at their peak treatment. Those with parkinsonian syndromes other than PD are typically excluded.

II.2.I. Tools for assessing Primary EndpointS
b. For Parkinsonism: UPDRS, total or Part III, or Parts II + III
c. For Dyskinesias: AIMS, Rush Dyskinesia Scale
d. Motor Fluctuations: Home diaries, Part IV of UPDRS

II.2.J. Specific criteria for early withdrawal and discontinuation
The investigator will discontinue a subject before completion of the study if consent is withdrawn, if the double-blind code is broken, if it is medically unacceptable to continue treatment due to adverse events or other reasons, or if pregnancy occurs.

II.2.K. Data analysis methods
The analysis of efficacy variables is based on the intention-to-treat strategy. All statistical tests are two-sided and p values <0.05 are considered statistically significant, though sometimes adjusted p-values are used to account for multiple comparisons. Many statistical methods can be used for the primary and secondary endpoints and include analysis of variance, analysis of covariance, and logistic regression.

II.3. Long-term studies
Long-term efficacy trials are usually nine months to five years.

II.3.A. Objectives
Studies are usually aimed at delaying the development of impairment, e.g. progressive parkinsonism, motor fluctuations or dyskinesias.

II.3.B. Primary Endpoints
For delay in progression of parkinsonism:
  a. Need to start dopaminergic therapy.
b. Prespecified increase in a standard measurement tool of parkinsonism (UPDRS)

For delay in development of motor fluctuations and dyskinesias
  a. Time to development of these complications
  b. % of the population at given time points who have the complication

II.3.C. Secondary endpoints
  a. For Delay in Parkinsonism Progression: secondary endpoints can include the primary endpoints for Delay in Development of Motor Fluctuations and Dyskinesias described above.
b. For Delay in Development of Motor Fluctuations and Dyskinesias, secondary endpoints can include the primary endpoints for Delay in Parkinsonism Progression described above
c. For all studies, other secondary endpoints are UPDRS scores, Hoehn and Yahr stage, Schwab and England rating scale score at specified time points.
d. For all studies, global secondary endpoints include Clinical Global Impression Severity and Clinical Global Impression Change scores as well as disease specific Quality of Life measures, such as the PDQ-39, and PDQUALIF or generic measures, such as the EuroQol and SF-36.
II.3.D. Exploratory endpoints
Depending on the chemistry of the intervention, exploratory endpoints can include non-motor elements of Parkinson’s disease such as depression, cognition, hallucinations, and dysautonomia. Rating scales for all of these are derived from standard scales in the medical literature and have not been developed specifically for Parkinson’s disease in most instances.

II.3.E. Study design
Randomized double-blind placebo controlled parallel studies are the gold standard. Studies usually involve the enrollment of several hundred patients and therefore multiple centers are usually involved. After screening and entry criteria are verified, subjects are randomized and are seen usually one month after study entry and thereafter on a three or six month schedule regularly. A final visit shortly after drug-exposure cessation is standard for safety monitoring and allows the detection of withdrawal effects on primary and secondary outcomes.

II.3.F. Study population
The study population depends on the question being addressed: Delay in Parkinsonism: In early disease, the subjects are usually on no treatment for parkinsonian signs at the time of enrollment and receive monotherapy with either placebo or study drug. In moderate disease, patients are enrolled who have inadequate efficacy from their current drug therapy and can often be on levodopa or another agent with a pharmacological mechanism that is different from the study drug under question. Because these studies are long in duration, some protocols permit addition of levodopa or other drugs with continuation in the study even after the primary endpoint is reached (e.g. need for starting dopaminergic therapy) so that secondary endpoints can still be measured.

Delay in Motor complications: These studies enroll patients who at baseline are in need of dopaminergic therapy. The study randomizes patients to standard dopaminergic therapy, usually levodopa, or the new study drug. Because these studies are long in duration, some protocols permit addition of additional levodopa in both groups if inadequate efficacy of treatment is encountered in the midst of the study period.

II.3.G. Specific Inclusion Criteria
Parkinson’s disease is defined clinically with diagnostic standards, such as UK Brain Bank criteria. Other parkinsonian syndromes that are not PD are not intentionally included.

a. Delay in Progression of Parkinsonism: Inclusion criteria for long-term studies are primarily based on short duration of PD (less than five years of symptoms or diagnosis), Hoehn and Yahr stage (usually less than Stage III, meaning no significant postural reflex compromise) and medication exposure (no dopaminergic medication and often other requirements such as no Coenzyme Q, no antidepressants).

b. Delay in Development of motor complications: Inclusion criteria for long-term studies of this type enroll PD patients starting dopaminergic therapy because of need to treat the symptoms of PD. Such patients must be Stage I-III (unilateral or bilateral disease that may include mild balance problems but no spontaneous falling). Because the studies examine the onset time to motor complications, patients must not have any of these signs at baseline.

c. For both types of studies, neuroimaging data are often an intrinsic part of the protocol, so patients must be able and willing to undergo these scans, and must not have claustrophobia or other limits that preclude participation in these tests.

II.3.H. Specific Exclusion Criteria
In each trial, patients cannot have an allergy to the product being tested. Past medication exposure, especially to levodopa, may exclude subjects. Because dopamine is a precursor to melanin, some studies
exclude patients with a past history of melanoma. Those with parkinsonian syndromes other than PD are typically excluded. Patients with another illness sufficiently severe as to jeopardize participation in the study or interpretation of results are excluded. Patients who fail screening tests for dementia and depression are typically excluded.

II.3.I. Tools for assessing Primary Endpoints
a. For Delay in Parkinsonism: Clinician’s assessment of the necessity to start dopaminergic therapy (for patient safety, job security, or quality of life), UPDRS, total or Part III, or Parts II + III
b. For Dyskinesias: Onset of first dyskinesias as assessed by diaries or by Part IV of the UDPRS, AIMS, Rush Dyskinesia Scale
c. Motor Fluctuations: Onset of first OFF period by Home diaries or by Part IV of UPDRS

II.3.J. Specific criteria for early withdrawal and discontinuation
The investigator will discontinue a subject before completion of the study if consent is withdrawn, if the double-blind code is broken, if it is medically unacceptable to continue treatment due to adverse events or other reasons, or if pregnancy occurs.

II.3.K. Data analysis methods
The analyses are usually based on the intention-to-treat strategy. All statistical tests are two-sides and p values <0.05 are considered statistically significant. The primary analysis for studies involving Delay in the Development of Clinical Progression of Parkinsonism or Motor Complications evaluate survival and calculate cumulative probability of reaching each end point. Differences in outcome are expressed as hazard functions, usually with the Kaplan-Meier estimate, which can be displayed graphically. It is important to censor individuals who either drop out or have limited follow-up. Hypothesis testing between groups can use the log-rank test or Cox’s proportional hazard regression modeling, which allows adjustment for multiple covariates.

III. PHASE III STUDIES FOR REGISTRATION OF NEW DRUGS

III.1. Outline a typical development plan
Phase III trials are conducted after initial demonstration of safety and efficacy of a drug. These trials investigate larger numbers of patients, to obtain further information on efficacy and safety. A primary goal of the Phase III development plan is to obtain the information necessary for product registration. Typical studies extend the observations from Phase II studies, often employing similar study designs, but with larger numbers of subjects and longer periods of observation. Because Parkinson’s disease is relatively uncommon, Phase III studies must invariably involve multiple centers in order to allow timely accrual.

Phase III trials will address the same two primary impairments – parkinsonism, or motor complications associated with chronic treatment (dyskinesias, motor fluctuations). The aim can be either to delay or prevent the development or slow the progression of the impairment or to provide symptomatic relief.

III.2 Long-term studies to slow or halt progression of parkinsonism
Conclusive demonstration that a drug can stop or slow progression of parkinsonism remains a challenge. The design of a trial with this objective must take into account some limitations. First, there is no biomarker of Parkinson’s disease progression. Primary endpoints are based on clinical measures of parkinsonism. Because most people with Parkinson’s disease require symptomatic therapy within several months to a few years of diagnosis, such studies typically enroll only those early in the course of disease who do not require symptomatic therapy. This design avoids the confounding effect of symptomatic
treatment on the efficacy endpoint measures. No alternative marker of disease has yet been accepted as a primary endpoint to assess progression of parkinsonism.

Demonstration of efficacy in slowing or stopping progression in those with more advanced disease, who require symptomatic therapy, presents even greater challenges. Few studies have attempted to demonstrate efficacy in this population. In those with more advanced parkinsonism, clinical measures of parkinsonism are confounded by the effects of symptomatic therapy, but withdrawal of the therapy is potentially harmful. Follow-up of those with more advanced disease until death or severe disability may pose practical difficulties. Some countries are fortunate in enabling research study participants to be prospectively “flagged” on a central database (e.g. National Health Service Central Register, UK) that automatically informs researchers when a participant has died and supplies a copy of the death certificate. A typical development plan to investigate a drug proposed to slow the progression of parkinsonism will include several multicenter, double blind, placebo-controlled trials investigating the safety and efficacy of one or several doses of the drug. Monotherapy trials are the norm. Patients are recently diagnosed and not in need of symptomatic therapy. Prior exposure to symptomatic antiparkinsonian therapy is precluded, or limited to a short time, to avoid confounding. Other agents proposed to slow the progression of parkinsonism are excluded. Because this type of drug would be expected to be used for many years, or even for the entire duration of disease, safety monitoring should allow detection of events expected to occur at moderate frequency (0.5 – 5%). In general, around 1200-1500 persons exposed in short and long term studies should be adequate to detect differences of around 2 to 3% across groups. To detect adverse events occurring after prolonged use of the drug, duration of monitoring should be at least 12 months for some subjects. Depending on the specific agent, special safety monitoring may be indicated.

### III.2.A. Objectives
Study goals typically are to slow or stop progression of parkinsonism

### III.2.B. Primary Endpoints

**Efficacy:**
- Need to start symptomatic antiparkinsonian therapy; or
- A prespecified worsening in a standard clinical assessment instrument, usually the UPDRS

**Safety:**
- All adverse events;
- Depending on the specific agent, monitoring special safety endpoints may be indicated. For example, agents thought to block apoptotic neuronal cell death may conceivably also present an increased risk of neoplasm, and special monitoring procedures for cancer may be appropriate.

### III.2.C. Secondary Endpoints

- Clinical assessment measures of parkinsonism, including UPDRS (if not a primary endpoint measure), Hoehn and Yahr stage, Schwab and England;
- Global measures such as need to start symptomatic therapy (if not a primary endpoint), Clinical Global Impression Severity or Change, and disease-specific Quality of Life measures, such as the PDQ-39 or PDQUALIF or generic measures such as EuroQol and SF-36;
- Neuroimaging outcomes measuring uptake of ligands specific to the dopamine system, such as \([^{123}\text{I}]\beta\text{-CIT}\) (2β-carbomethoxy-3β-[4-iodophenyl]) and single photon positron emission tomography (SPECT) or \([^{18}\text{F}]\)-dopa and positron emission tomography (PET) scanning indices.

### III.2.D. Exploratory endpoints
- Endpoints targeting nonmotor features of Parkinson’s disease, such as cognition, dysautonomia, depression;
- Endpoints investigating response to symptomatic therapy once initiated;
c. Endpoints investigating development of the complications of chronic dopaminergic therapy (dyskinesias, fluctuations, hallucinations or psychosis)

d. Although death is a logical endpoint when investigating agents thought to alter the course of disease, the long disease duration (10 or more years on average, depending on the age at onset), makes death an impractical outcome for many drug development plans.

III.2.E. Study Design

Randomized double-blind placebo controlled parallel studies are the gold standard. Individual studies generally involve the enrollment of at least 300 patients and therefore multiple centers are needed.

Because there is inevitable subjectivity in endpoint determination, it is almost always desirable to require that the primary outcome measure be determined by the same rater, at a minimum for key time points (such as enrollment and endpoint). It may be desirable to identify specific raters within a center and/or to specify a required level of expertise with the primary efficacy measure. A blocked randomization, either by investigator or by center, is another approach to minimize the effect of between-rater variability in endpoint determination.

To avoid “unblinding” and the potential for biased endpoint assessment, two raters may be used -- a “treating” investigator who evaluates the patient at each visit, and a “blinded” investigator who determines performance on study primary outcome measures only at key visits, and is otherwise prohibited from knowledge of the subject. The use of video-assessment enables a core group of central raters to assess patients across a wide geographical distribution, but may be problematic for some impairments e.g. rigidity.

Subjects are typically assessed 1 – 6 weeks after the initiation of study drug, depending on the safety and pharmacology of the specific agent. Subsequent visits typically occur at 3-6 month intervals. Telephone follow up for safety monitoring may be planned between in-person visits. The duration of exposure for any one individual will vary depending on the proposed mechanism of the drug under development, but a minimum of 9 months follow up is though necessary to demonstrate efficacy in slowing the progression of parkinsonism.

One or more visits after drug-exposure cessation is standard for safety monitoring. When the drug has known or suspected symptomatic benefit in parkinsonism, the primary outcome measurement may be most easily interpreted only after study drug has been withdrawn. The symptomatic benefit may be mild, and only identifiable when symptoms worsen after drug withdrawal. The primary efficacy outcome may at times be determined at an interval after withdrawal of the drug under development. When planning the timing of post-treatment assessments, it will be important to consider the pharmacology of the drug under study, so that the study drug is washed out when assessments are performed.

Follow up is often continued after primary efficacy data have been obtained. Extended follow up can be especially valuable in monitoring safety, and to assess secondary outcomes such as the development of dyskinesia, motor fluctuations or psychosis.

III.2.F. Study population

Subjects typically have recently diagnosed Parkinson’s disease, not requiring symptomatic therapy and with little or no prior exposure to symptomatic or proposed neuroprotective therapies. Because Parkinson’s disease is diagnosed only by clinical criteria, and the full complement of diagnostic signs may not manifest for several years, it is expected that a percentage of those meeting diagnostic criteria with recently diagnosed Parkinson’s disease will be misclassified. Experts with greater familiarity with Parkinson’s disease have greater long-term diagnostic accuracy, but some error is inevitable. The greater potential for misclassification in early disease should be considered in determining sample size. Whilst misclassification may result in attenuated effect estimates, assuming no therapeutic benefit for the patients
who have been misdiagnosed, this result may be a more realistic estimate of treatment benefits outside trials, where expert diagnosis may not always be available prior to initiating therapy.

III.2.G. Specific Inclusion Criteria

a. Parkinson’s disease is defined clinically using published diagnostic standards, such as the UK Brain Bank and NIH criteria. These criteria exclude those with signs suggesting other parkinsonian syndromes.

b. Disease duration is typically less than five years after diagnosis. Symptom duration may be used, although this measure is dependent on patient report and may be less reliable.

c. Symptomatic treatment must not be needed in the opinion of the subject and the investigator. Enrolling subjects would need to be comfortable without symptomatic therapy if parkinsonism did not progress.

d. Mild disease severity defined using clinical criteria, such as Hoehn and Yahr stage less than III. Other clinical measures such as tremor scores or Schwab and England scores may also be used.

e. No or limited prior treatment with drugs proposed to slow disease progression (e.g., selegiline, coenzyme Q10).

f. No or limited prior treatment with symptomatic antiparkinsonian treatments.

g. If imaging is a secondary outcome, patients must be able and willing to undergo these scans, and must not have claustrophobia or other limits that preclude participation in these tests. In many cases, only a subgroup of subjects participate in the imaging study.

III.2.H. Specific Exclusion Criteria

a. Current treatment with drugs that could alleviate parkinsonism (e.g., dopaminergic, anticholinergic drugs). In early Parkinson’s disease, the appropriate washout period for the available symptomatic therapies is not well established. Generally a washout period of at least 4 weeks is desirable, in order to avoid excess early terminations due to enrollment of subjects not able to function without antiparkinsonian treatment.

b. Current treatment with drugs that could worsen (e.g., most antipsychotics, some antiemetics) parkinsonism

c. Any serious illness that may affect participation

d. Known allergy to study drug or related compounds

e. Use of medication thought to interact with study drug

f. At risk for an adverse effect of a specific drug

III.2.I. Tools for assessing Primary Endpoints

The endpoints are determined by the investigator using clinical skills. Familiarity with the disease and the specific instruments used is therefore critical to the integrity of the study endpoint. Training in the use of the endpoint instruments is desirable.

III.2.J. Specific criteria for early withdrawal and discontinuation

The investigator will discontinue a subject before completion of the study if consent is withdrawn, if the double-blind code is broken, if it is medically unacceptable to continue treatment due to adverse events or other reasons, or if pregnancy occurs. A special case occurs when the primary endpoint measure is the investigator-determined need for symptomatic therapy. If the subject initiates symptomatic therapy prior to the investigator-determined end point, it may be preferable to continue to observe the subject on study drug and symptomatic therapy when possible, in order to obtain additional safety information.

III.2.K. Data analysis methods

The primary analyses are based on the intention-to-treat strategy. All statistical tests are two-sided. Generally p values <0.05 are considered statistically significant. The primary analysis approach typically evaluates survival and calculates cumulative probability of reaching each end point. Differences in
outcome are expressed as hazard functions, usually with the Kaplan-Meier estimate, which can be displayed graphically. It is important to censor individuals who either drop out or have limited follow-up. Hypothesis testing between groups can use the log-rank test or Cox’s proportional hazard regression modeling, which allows adjustment for multiple covariates.

III.3 Long term studies to provide symptomatic improvement in parkinsonism

While there are a number of agents with demonstrated antiparkinsonian efficacy, none provides sustained symptomatic benefit throughout the course of this lifelong disorder. The acute and chronic side effects of established therapies are additional sources of concern. Trials of new therapies should be designed to address these concerns, with the goal of developing new drugs with more favorable efficacy and side effect profiles.

A typical development plan to investigate a drug proposed to provide symptomatic improvement of parkinsonism will include several multicenter, double blind, placebo-controlled trials investigating the safety and efficacy of one or several doses of the drug. Monotherapy trials, comparing the study drug to placebo, will in most cases be limited to early disease, enrolling “de novo” patients receiving symptomatic therapy for the first time. In advanced Parkinson’s disease, the study-drug will typically be given as adjunctive therapy, and compared to placebo given adjunctively, as it would be unethical to withdraw existing therapies. Most commonly, the efficacy of the new agent when given in combination with a dopaminergic agent (usually L-dopa plus decarboxylase inhibitor) is compared to the efficacy of placebo combined with the same agent. The addition of an adjunctive antiparkinsonian agent can result not only in improvement of parkinsonism, but also in the new onset of dopaminergic side effects, or the worsening of existing side effects, such as dyskinesias or psychosis. Specific monitoring for this possibility, and provisions for adjustment of therapies, appropriate to the specific drugs, should be included in the study protocol. In addition a global measure such as a disease specific quality of life measure is essential as it is otherwise impossible to interpret an improvement in motor function coupled with a deterioration in side effects. An alternative design compares the new drug to standard therapy. Design of such studies is often difficult due to uncertainty regarding equivalence of dosage. Some regulatory agencies may be less receptive to comparison study designs. As for all development plans, close contact with scientists in the regulatory agencies is essential.

Safety evaluations should take into account the chronic use expected for most drugs in this category. Therefore, safety monitoring should allow detection of events expected to occur at moderate frequency (0.5 – 5%). In general, around 1200-1500 persons exposed in short and long term studies should be adequate to detect differences of around 2 to 3% across groups. To detect adverse events occurring after prolonged use of the drug, duration of monitoring should be at least 12 months for some subjects. Depending on the specific agent, special safety monitoring may be indicated.

III.3.A. Objectives

Studies are aimed at treating impairment due to parkinsonism or improving existing dyskinesias or motor fluctuations.

III.3.B. Primary Endpoints

For parkinsonism:

a. Comparison of the Unified Parkinson’s Disease Rating Scale (total score) relative to baseline scores.

b. Comparison of the UPDRS Motor Examination (Part III) can be used as well, or the combined Activities of Daily Living and Motor Examination score (Parts II and III) relative to baseline scores. The UPDRS Part I includes nonmotor features and does not distinguish between primary features of disease and drug-induced side effects. For this reason, some prefer not to use Part I when assessing a new drug as adjunctive therapy along with a dopaminergic agent.

c. The UPDRS is internationally utilized and has largely replaced earlier scales like the Columbia and Webster scales.
The Hoehn and Yahr scale was formerly used, but it is a non-continuous scale, poorly responsive to interventions, and therefore more frequently used currently to describe patient groups and define entry criteria, rather than serving as a primary end-point.

For motor fluctuations:

a. Dyskinesias are usually rated with the Abnormal Involuntary Movement Scale (AIMS), or the Rush Dyskinesia Scale. Dyskinesias are often intermittent, and an at-home diary may be used. However, a self-report diary will likely identify only dyskinesias of moderate to severe intensity, as mild dyskinesias may be missed by the patient experiencing them.

b. Motor fluctuations are measured with at-home diaries for which patients undergo training in the study center on the operational definitions of “ON” (good medication response), “ON with disabling dyskinesias” (good medication response, but with superimposed involuntary movements that interfere with activities), and “OFF” (poor medication response). Decrease in overall OFF time without an increase in ON with dyskinesias indicate improved motor fluctuations.

c. Global measures on motor fluctuations and dyskinesias can be obtained from UPDRS Part IV.

III.3.C. Secondary endpoints

For parkinsonism:

a. Hoehn and Yahr stage, the Schwab and England rating scale

b. Dyskinesias and motor fluctuation as secondary outcomes are measured as described in Primary endpoints

c. Global scales like the Clinical Global Impression Severity and Clinical Global Impression Change scores are also used, as well as disease specific Quality of Life measures, such as the PDQ-39, and PDQUALIF and generic measures such as the EuroQol and SF-36.

For motor fluctuations or dyskinesias:

a. Secondary endpoints include UPDRS and all primary endpoints listed under Parkinsonism.

III.3.D. Exploratory endpoints

Depending on the chemistry of the intervention, exploratory endpoints can include non-motor elements of Parkinson’s disease such as depression, cognition, somnolence, hallucinations or dysautonomia. Rating scales for all of these are derived from standard scales in the medical literature. In most cases these have not been developed specifically for Parkinson’s disease.

III.3.E. Study Design

Randomized, double-blind, placebo-controlled, parallel group studies are the standard. After screening and entry criteria are verified, subjects are randomized and are seen regularly during study-drug intervention and then withdrawn from the drug and seen at a close-out visit. Some studies assign patients to one or more fixed doses or placebo. Other designs allow dose titration to an efficacy endpoint (e.g., loss of motor fluctuations) or to a pre-determined maximum, if tolerated. Visit frequency is determined in part by the pharmacologic and safety profile of the study drug, and the endpoint(s) of interest. Weekly or biweekly visits are typical during the titration phase, followed by longer between visit intervals, such as 4 – 12 weeks. There is an end-of-exposure visit and a final visit one week after drug-exposure cessation.

Because there is inevitable subjectivity in endpoint determination, it is almost always desirable to require that the primary outcome measure be determined by the same rater, at a minimum for key time points (such as enrollment and endpoint). It may be desirable to identify specific raters within a center and/or to specify a required level of expertise with the primary efficacy measure. A blocked randomization, either by investigator or by center, is another approach to minimize the effect of between-rater variability in endpoint determination.
To avoid “unblinding” and the potential for biased end point assessment, two raters may be used -- a “treating” investigator who evaluates the patient at each visit, and a “blinded” investigator who determines performance on study primary outcome measures only at key visits, and is otherwise prohibited from knowledge of the subject. The use of video-assessment enables a core group of central raters to assess patients across a wide geographical distribution, but may be problematic for some impairments e.g. rigidity.

Follow up is often continued after primary efficacy data have been obtained. Extended followup can be especially valuable in monitoring safety, and to assess chronic efficacy.

III.3.F. Study population
The study population depends on the question being addressed:
Parkinsonism: In early disease, the subjects are usually on no treatment for parkinsonian signs at the time of enrollment and receive monotherapy with either placebo or study drug. In moderate and advanced disease, patients are enrolled who have inadequate efficacy from their current drug therapy. Current therapy is typically levodopa or another agent with a pharmacological mechanism that is different from the study drug under question. When developing a drug for adjunctive use, the determination of what standard therapies will be acceptable must be made. Most commonly new adjunctive treatments are compared to placebo in patients receiving levodopa.

Motor complications: In advanced disease where the focus is usually on motor complications, patients must have dyskinesias and/or motor fluctuations of sufficient severity to warrant intervention.

Patients with another illness sufficiently severe as to jeopardize participation in the study or interpretation of results are excluded. Patients who fail screening tests for dementia and depression are typically excluded.

III.3.G. Specific Inclusion Criteria
Parkinson’s disease is defined clinically based on various diagnostic standards, such as UK Brain Bank criteria. Inclusion criteria for admitting mild, moderate or advanced patients with PD are primarily based on Hoehn and Yahr stage and medication exposure. Early monotherapy studies restrict subjects to Stage I-III (unilateral or bilateral disease that may include mild balance problems but no spontaneous falling) and sometimes to Stage I-II only (no balance problems). Studies of drugs that are added to standard treatment in moderate Parkinson’s disease usually restrict patients to Stage II-IV, and therefore include patients with poor balance. For studies of motor complications, inclusion criteria usually require baseline scores for the target problem sufficiently severe enough to allow change determination during the trial. For dyskinesia, a minimal baseline score on the AIMS (variably 7-10) is often used, and for motor fluctuations inclusion often requires a minimal 25% or more OFF time on diaries or the UPDRS Part IV for study entry. These criteria are introduced to avoid “floor effects”. The existing antiparkinsonian drug regimen should be optimized before determining eligibility.

III.3.H. Specific Exclusion Criteria
In each trial, patients cannot have an allergy to the product being tested. Current exposure to various medications and past exposure to levodopa may exclude subjects from the early monotherapy trials. Because dopamine is a precursor to melanin, studies of drugs that alter levodopa bioavailability or metabolism exclude patients with a past history of melanoma. Almost no intervention studies permit Stage V patients who are wheelchair-bound at their peak treatment. Those with parkinsonian syndromes other than PD are typically excluded.

III.3.I. Tools for assessing Primary Endpoints
b. For Parkinsonism: UPDRS, total or Part III, or Parts II + III
c. For Dyskinesias: AIMS, Rush Dyskinesia Scale
d. Motor Fluctuations: Home diaries, Part IV of UPDRS
III.3.J. Specific criteria for early withdrawal and discontinuation
The investigator will discontinue a subject before completion of the study if consent is withdrawn, if the double-blind code is broken, if it is medically unacceptable to continue treatment due to adverse events or other reasons, or if pregnancy occurs.

III.3.K. Data analysis methods
The analysis of efficacy variables is based on the intention-to-treat strategy. All statistical tests are two-sides and p values <0.05 are considered statistically significant. Many statistical methods can be used for the primary and secondary endpoints and include analysis of variance, analysis of covariance, and logistical regression.

III.4. Long term studies to delay the development of dyskinesias or motor fluctuations
Dyskinesias and motor fluctuations are inevitable side effects for most patients requiring levodopa. These side effects are much less commonly associated with other antiparkinsonian agents. However, the majority of persons with Parkinson’s disease eventually require levodopa therapy.

III.4.A. Objectives
Studies usually are aimed at delaying the development of motor fluctuations, dyskinesias or both.

III.4.B. Primary Endpoints
a. Time to development of these complications
b. % of the population at given time points who have the complication

III.4.C. Secondary endpoints
a. Measures of parkinsonian impairment, such as UPDRS scores, Hoehn and Yahr stage, Schwab and England rating scale
b. Clinical Global Impression Severity and Clinical Global Impression Change
c. Quality of Life measures, such as the PDQ-39, and PDQUALIF.
d. Neuroimaging outcomes measuring uptake of ligands specific to the dopamine system, such as $^{[123]I}$ß-CIT (2ß-carbomethoxy-3ß-[4-iodophenyl]) and single photon positron emission tomography (SPECT) or $^{[18F]}$-dopa and positron emission tomography (PET) scanning indices.

III.4.D. Exploratory endpoints
Depending on the chemistry of the intervention, exploratory endpoints can include non-motor elements of Parkinson’s disease such as depression, cognition, hallucinations, and dysautonomia. Rating scales for all of these are derived from standard scales in the medical literature and have not been developed specifically for Parkinson’s disease in most instances.

III.4.E. Study Design
 Randomized double-blind placebo controlled parallel studies are the standard. Studies usually involve the enrollment of several hundred patients and therefore multiple centers are usually involved. After screening and entry criteria are verified, subjects are randomized and are seen usually one month after study entry and thereafter at three or six month intervals. A final visit shortly after drug-exposure cessation is standard for safety monitoring and allows the detection of withdrawal effects on primary and secondary outcomes.

III.4.F. Study population
These studies enroll patients who at baseline are in need of symptomatic antiparkinsonian therapy. The study randomizes patients to standard dopaminergic therapy, usually levodopa, or the new study drug. Because these studies are long in duration, some protocols permit addition of additional levodopa in both groups if inadequate efficacy of treatment is encountered in the midst of the study period.
III.4.G. Specific Inclusion Criteria
Parkinson’s disease is defined clinically with diagnostic standards, including UK Brain Bank criteria. Other parkinsonian syndromes that are not PD are excluded. PD patients must be newly in need of symptomatic antiparkinsonian therapy, typically Hoehn and Yahr Stage II or III. Patients should have no prior exposure or very minimal prior exposure to dopaminergic drugs and should not have motor fluctuations or dyskinesias.

If neuroimaging endpoints are important to the study design, patients must be able and willing to undergo these scans, and must not have claustrophobia or other limitations that preclude participation in these tests.

III.4.H. Specific Exclusion Criteria
In each trial, patients cannot have an allergy to the product being tested. Past medication exposure, especially to levodopa, may exclude subjects. Because dopamine is a precursor to melanin, some studies exclude patients with a past history of melanoma. Parkinsonian patients who carry other diagnoses besides Parkinson’s disease are excluded.

III.4.I. Tools for assessing Primary Endpoints
a. For Dyskinesias: Onset of first dyskinesias as assessed by diaries or by Part IV of the UPDRS, AIMS, Rush Dyskinesia Scale
b. Motor Fluctuations: Onset of first OFF period by Home diaries or by Part IV of UPDRS

III.4.J. Specific criteria for early withdrawal and discontinuation
The investigator will discontinue a subject before completion of the study if consent is withdrawn, if the double-blind code is broken, if it is medically unacceptable to continue treatment due to adverse events or other reasons, or if pregnancy occurs.

III.4.K. Data analysis methods
The analyses are based on the intention-to-treat strategy. All statistical tests are two-sided and p values <0.05 are considered statistically significant, though sometimes adjusted p-values are used to account for multiple comparisons. The primary analysis for studies involving Delay in the Development of Motor Complications evaluates survival and calculates cumulative probability of reaching each end point. Differences in outcome are expressed as hazard functions, usually with the Kaplan-Meier estimate, which can be displayed graphically. It is important to censor individuals who either drop out or have limited follow-up. Hypothesis testing between groups can use the log-rank test or Cox’s proportional hazard regression modeling, which allows adjustment for multiple covariates.

IV. OTHER STUDIES (NEW INDICATION TRIALS, PRAGMATIC TRIALS)

IV.1 Special clinical problems
Outside of the primary motor elements of PD (parkinsonism and motor complications), PD patients experience a number of other disabilities, including hallucinations, dementia, depression, dysautonomia, sexual dysfunction, and fatigue. Drugs that are useful in for treating these symptoms in other medical contexts can be tested in PD through randomized double-blind placebo-controlled trials of PD subjects with the target problem. The example of hallucinations is provided, as a prototype, because it has been studied more than the other special clinical problems listed. For each of the other conditions, similar studies can be performed using PD patients with the target problem and appropriately designed measurement tools adapted from other medical fields.
IV.2. Hallucinations

IV.2.A. Objectives
Reduce the frequency or severity of hallucinations in drug-treated patients with chronic PD and hallucinations.

IV.2.B. Primary Endpoints
Change scores on standardized measures of hallucinations or global psychiatric disturbance.

IV.2.C. Secondary endpoints
Because drugs that improve hallucinations generally block dopamine receptors, the risk of aggravating PD is substantive and therefore secondary endpoints include standard assessments of parkinsonism, including UPDRS and Hoehn and Yahr stage.

IV.2.D. Exploratory endpoints
Scores on inventories for Depression, Cognition, and Quality of Life.

IV.2.E. Study Design
Open label exploratory and double blind placebo-controlled or clozapine-controlled trials have been conducted. These studies are usually short-term (4 weeks to 3 months), and parallel in design.

IV.2.F. Patient sample
Subjects with chronic hallucinations, with severity and frequency defined by clinical judgment as in need of treatment or by specific scores on screening tests like the Hallucination and Delusion items of the Neuropsychiatric Inventory, are enrolled in studies of agents shown to be useful against hallucinations in psychiatric populations, usually schizophrenia. Traditionally, drug dosage ranges in PD are up to 100 times less than in schizophrenia.

IV.2.G. Specific Inclusion Criteria
Parkinson’s disease is defined clinically based on various diagnostic standards, such as UK Brain Bank criteria. Inclusion criteria must establish that hallucinations began after chronic exposure to dopaminergic drugs in order to exclude the contamination of the sample by subjects with Dementia with Lewy Bodies. Entry criteria must establish that hallucinations are frequent and severe enough at baseline to warrant intervention and that scores on the baseline hallucination assessment are high enough to permit detection of change with the intervention.

IV.2.H. Specific Exclusion Criteria
In each trial, patients cannot have an allergy to the product being tested. Current exposure to other treatments for hallucinations is not permitted and usually an abstinence from such drugs for a minimum of 4 weeks is required. Almost no intervention studies permit Stage V patients who are wheelchair-bound at their peak treatment. Parkinsonian patients who carry other diagnoses besides Parkinson’s disease are excluded.

IV.2.I. Tools for assessing Primary Endpoints
Specific hallucination scales includes the Scale for Positive Symptoms (SAPS), the Parkinson Psychosis Scale, Item I (Thought Disorder) of Part I of the UPDRS, and individual items on various scales including the Neuropsychiatric Inventory. Global scales include the Brief Psychiatric Rating Scale, total Part I score of the UPDRS, and the Clinical Global Impressions scale.

IV.2.J. Specific criteria for early withdrawal and discontinuation
The investigator will discontinue a subject before completion of the study if consent is withdrawn, if the double-blind code is broken, if it is medically unacceptable to continue treatment due to adverse events or other reasons, or if pregnancy occurs.
IV.2.K. Data analysis methods
The analysis of efficacy variables is based on the intention-to-treat strategy. All statistical tests are two-sided and p values <0.05 are considered statistically significant. Multiple statistical methods can be used for the primary and secondary endpoints and include analysis of variance, analysis of covariance, and logistic regression. Clozapine is the only drug that has been shown to have efficacy for hallucinations in double-blind placebo-controlled trials but it is expensive and has potentially dangerous side effects. As it is unlikely that a new drug will be superior to clozapine but may be cheaper and/or safer, a study of a new agent against clozapine could be designed as an equivalence trial as long as it is powered to detect sufficiently narrow confidence intervals around equivalence. It is generally the case that equivalence trials require a larger sample size for the same power than a conventional trial.

IV.2. Surgical interventions: Deep brain stimulation, lesions, and cellular replacement therapies
The increased knowledge of the disrupted anatomical circuitry in PD has prompted laboratory studies and clinical trials of surgical interventions. On the premise that the degeneration in PD leaves several nuclei overactive from loss of inhibition, lesions and high voltage electrical stimulation have been applied to several deep brain structures, including the globus pallidum, thalamus and subthalamic nucleus. Alternatively, cellular replacement therapies focus on transplanting dopaminergic cells into the striatum in an attempt to reinnervate denervated structures. These studies are designed to evaluate both short-term and long-term (one to two years) efficacy, having the same objectives, rating tools and analytic methods described above for the treatment of parkinsonism and motor complications. Only special issues that are particular to these surgical trials are listed below:

IV.2.A. Study Design
Whereas randomized double-blind controlled parallel studies are the standard design in medication trials, this model is more difficult to effect in surgical trials. Lesion studies have primarily been open-label observations, and only a few have used a prospectively followed comparison group that receives optimal medical care. A few have randomized patients between two different surgical procedures. For deep brain stimulation, blinded ratings with the stimulator turned on and turned off have been used for comparisons. In cellular replacement therapies, the randomized, double-blind placebo-controlled parallel design that is typical of medical trials has been most closely replicated. In these cases, subjects are randomized between treatment groups and those who are assigned the control group go to the operating room, have a skull burr hole placed, but no needle penetration or cellular placement into the brain occurs. The surgical investigator is the only person on the research team who is unblinded, and all ratings are performed by investigators who were not involved in the surgery. In all studies, subjects are evaluated at baseline (often with more than one baseline assessment) and then seen regularly after surgery, usually at one month, and every three months thereafter during the trial. The score at the final visit usually serves as the primary outcome measure having adjusted for baseline scores. Often the percentage change in baseline scores is presented across different interventions. A common feature of all these trials is that the sample size is much smaller given the complexity and expense of the interventions. This makes the use of standardized outcome measures even more important as, inevitably, pooling results through the use of meta-analysis will be required to reduce the likelihood of both type I and type II errors. One special feature of surgical trials is the ability to assess an intervention undertaken either unilaterally or bilaterally. If subjects are randomised to have a different procedure for each side then, they can act as their own controls and matched methods of analysis are required as in cross-over studies. More typically comparisons are made by side so that 20 patients treated bilaterally provide 40 outcome measures. In this case, it is important to remember that each observation is not truly independent as they clustered within individuals and more complex statistical methods are required to allow for this clustering.
IV.2.B. Study population
The study population for surgical interventions are subjects with advanced PD who have failed other therapies, but still show an objective improvement (even if for short intervals) to dopaminergic therapy. They have motor complications in the form of dyskinesias and/or motor fluctuations. During “on” periods, they must be Hoehn and Yahr Stage I-III and during “off” periods, they must be Stage III-V.

IV.2.C. Specific Inclusion Criteria
Because surgical intervention is a major medical treatment, patients must be in good health other than their PD. Most studies require good cognitive status (MMSE usually at least 24) and no hallucinations. They must clearly understand all surgical risks and have a caregiver who can participate in the program.

IV.2.D. Specific Exclusion Criteria
Parkinsonian patients who carry other diagnoses besides Parkinson’s disease are excluded. Dementia, hallucinations and other significant behavioral problems usually exclude patients from these trials.

V. SPECIAL CONSIDERATIONS
Placebo effects are frequent and substantive in PD trials. The dopamine system is directly involved in the regulation of reward mechanisms, expectation, motivation and vigilance. Positron emission tomography [11C]-raclopride binding studies document evidence of increased striatal dopamine release in PD subjects responding to placebo treatment. Because most drugs or interventions being studied in PD share dopaminergic augmentation mechanisms, separating primary dopaminergic effects due to the intervention vs. dopaminergic effects due to study participation (placebo effects) must be delineated. Whereas a positive effect on parkinsonism is anticipated in both the control and study group in PD, the outcome scores, after adjusting for baseline scores, must be significantly better in the intervention arm than in the placebo-treated arm before one can conclude that improvement is due to the intervention.

VI. EXAMPLES OF LANDMARK TRIALS
Treatment of parkinsonism

Treatment of motor complications

Prevention of clinical progression of parkinsonism

Prevention of disease mortality
Prevention of development of motor complications

Treatment of special issues

Surgical interventions

VII. SUGGESTED READINGS
3. Tanner CM, PD or not PD?—that is the question. Neurology 2003;61:5-6.
Chapter 20. Multiple Sclerosis and Other Demyelinating Diseases

Giancarlo Comi, M.D. ¹

Michel Clanet, M.D. ²

Fred Lublin, M.D. ³

Vittorio Martinelli, M.D. ⁴

Chris H. Polman, M.D. ⁵

Per Solberg Sorensen, M.D. ²

¹ Department of Neurology
University Vita-Salute IRCCS San Raffaele
Via Olgettina, 60
20132 Milan
ITALY
Tel +39/02/26432990
Fax +39/02/26433746
Email comi.giancarlo@hsr.it

² Department of Neurology
Hospital Purpan
Toulouse
FRANCE

³ Department of Neurology
VU Medical Centre
Amsterdam
THE NETHERLANDS

⁴ Department of Neurology
University Vita-Salute IRCCS
San Raffaele, Milan
ITALY

⁵ Department of Neurology
Copenhagen University Hospital
DENMARK

Corresponding Author:
Giancarlo Comi, M.D.
I. INTRODUCTORY REMARKS

Multiple Sclerosis (MS) is a devastating disease affecting predominantly the white matter of the central nervous system. It is the most frequent cause of disability in young adults, after car accidents. The disease affects about one million people, mostly in developed countries. In fact, the disease has a peculiar geographic distribution, probably as a result of an interaction of genetic and environmental factors, being more frequent in North America and North Europe and almost absent in equatorial regions and in Asia. Environmental factors are completely unknown and genetic factors have been only partially revealed, with a key role for genes involved in regulatory mechanisms of the immune system. No preventing strategies are available.

In 1993 FDI approved the use of Interferon Beta 1b for the treatment of relapsing remitting MS. Shortly later both Interferon beta 1a and Interferon beta 1b were approved by drug agencies worldwide and Glatiramer Acetate was approved in the end of the nineties. Finally Interferon beta 1b was approved also for treatment of secondary progressive multiple sclerosis. Soon after the approval of these immunomodulating agents, new phase II-III clinical trials were started to explore the best dose and frequency of injection for available therapies, to evaluate combination treatments and to test safety and efficacy of new treatments. Immunosuppressive agents have been extensively used to treat MS since more than thirty years, however only recently mitoxantrone has been approved by FDI for treatment of active MS patients not responding to immunomodulating agents.

There are converging and convincing evidences that early treatment is more effective in MS and that both Interferons (IFNs) and Glatiramer Acetate (GA) produce little or no benefits in the progressive phases of the disease. Multi-weekly injections, particularly in the initial period of treatment, resulted superior to the weekly injection. However increasing dose and frequency of injection result in a higher frequency of anti-interferon antibodies, which, if persistently present may limit the efficacy of treatment.

New therapies have a broad range of targets, including the T cell receptor, the co-stimulation molecules, the blood-brain barrier permeability, chemokines, etc. It has been estimated that about one hundred clinical trials are ongoing in Europe and North America. Most of these trials have a placebo control group, which raises ethical and practical issues because of the availability of active treatments. Ethical concerns for placebo arm are particularly relevant for phase III clinical trials because the study duration is usually not shorter than 2 years.

Combination therapy, an approach derived by treatment of tumours, is already in use in some MS centres, however the efficacy of this therapeutic strategy has not been proved yet. Two approaches are generally used: 1) the comparison between treatment A and the combined treatment A+B (where treatment A is Glatiramer Acetate or Interferon beta 1b); 2) treatment A is compared to treatment B followed by treatment A (where treatment B is usually an immunosuppressive agent).

MS is not a homogenous disease. Three main courses have been described, relapsing-remitting (RR), secondary progressive and primary progressive, to which we should add atypical variants such as Marburg’s disease, Devic’s disease and Balo’s disease. Post mortem and biopsy studies revealed that MS subtypes are characterised by different pathological findings and that in the same patient pathogenetic mechanisms may vary along the disease course. This intra- and inter-individual variability of the dysfunction of the immune system may be one of the reasons why the response to immunomodulators is so variable in MS patients. It is possible that future trials will concentrate on a more homogenous group of patients using a combination of clinical, instrumental and genetic parameters.

The efficacy of IFNs and GA on MS disease activity, at least in patients with RR disease is definitely proved. However, also in this population of patients, it is still debated if the “anti-inflammatory” activity of IFNs and GA also produce long term effects on disability. There are some contradictions among
clinical trials performed in RRMS on the effects on disability, explained by methodological problems (problems in the definition of progression, sample size, study duration, poor responsiveness of EDSS etc.) and by differences in the included population. Long term placebo controlled studies in RRMS, so necessary to prove trial benefits for the prevention of the accumulation of irreversible nervous damage, are not possible anymore for ethical and practical reasons. On the other hand, comparison with epidemiological data are very difficult to interpret. The open or blind extension of placebo controlled clinical trials with the comparison of patients with early and delayed active treatment may give problems of interpretation of the results because of the patients lost to the follow up and the frequency of patients switched to other treatments.

In Secondary Progressive MS all clinical trials failed to show efficacy and only the European Multicenter trial was able to demonstrate a significant effect of IFN beta1b on the proportion of patients with confirmed increase of EDSS score. No significant effects on disability were observed in clinical trials evaluating the effects of IFNs and GA in Primary Progressive MS.

The problem of paraclinical endpoints has been widely discussed in many meetings. Their use for phase II clinical trials in RR patients, has been accepted by national and supranational drug agencies because active MRI lesions are a more precise and sensitive measure of disease activity than the count of the number of relapses. In phase III trials in RRMS and in progressive MS courses only disability measures are accepted as primary endpoint; MRI surrogate markers may be used only as secondary or exploratory measures.

II. PHASE II STUDIES FOR REGISTRATION OF NEW DRUGS

Phase II studies in MS have been mostly short term, i.e., approximately six months and designed to answer two main questions: 1) proof of principle of the proposed agent and 2) dosing information. Safety monitoring, as always, is also a prime consideration. To date, most phase II MS clinical trials have involved immunomodulatory or immunosuppressive agents. We should soon see trials of neuroprotective agents. For the immunomodulating agents, the basis of the trials often comes from either animal models, such as experimental allergic encephalomyelitis, or from other therapeutic areas where there is a proposed autoimmune pathogenesis, such as rheumatoid arthritis or Crohn’s disease. This strategy centers on altering the immune response to a more suppressive bias, i.e., from Th1 to Th2 or from helper cytokines to suppressive cytokines. Such an approach has advantages over a more generalized immunosuppressive approach, but there are ongoing and planned immunosuppressive strategies, as well. Recent results suggest that such translation is not without some hazard and the need to assure that an agent does not worsen MS is paramount. The problem lies, in part, with our limited understanding of the full mechanisms underlying the production of flare ups and deficit production in MS, such that therapeutic approaches considered potentially beneficial have turned out to produce the opposite result, e.g., anti-TNF directed therapies. Therefore, safety assessments must not only look for adverse events from an experimental agent but also assess whether the agent might have a negative effect on disease course. This can be monitored by either the relapse rate (looking for an increase), change in level of disability (a more difficult outcome to monitor in phase II trials) or increase in MRI activity. The latter is the easiest to accomplish, for the reasons detailed below, and has been utilized as a primary outcome measure in safety trials of combination therapy in MS. The logical basis for this is that most current MS therapies produce a reduction in the number of gadolinium enhancing lesions. Therefore, if the addition of another immunomodulating agent produced an increase in the number of gadolinium enhancing lesions would indicate an adverse interaction between the two agents.

For relapsing-remitting MS, phase II trials most commonly measure a relative reduction in gadolinium enhancing lesion activity as the indication of efficacy. As gadolinium enhancing lesions occur 6-10x as frequently as clinical exacerbations, they provide greater power for detecting differences. This power,
through frequent scanning, provides for an ease of detection and a need for smaller numbers of patients than a clinical outcome would require. In such trials, relapse rate reduction is an important secondary outcome. As these trials usually are of six months duration, changes in EDSS are less likely to be seen. Use of gadolinium enhanced lesions is only reasonable when the putative mechanism of action of the tested agent is expected to impact on the blood-brain barrier (BBB). Such would be the case with most anti-inflammatory agents and adhesion molecule blockers. Agents that might act away from, or independent of the BBB or that might not affect inflammation should have outcome measures that either reflect a clinical change, such as relapse rate, or use MRI metrics that relate to tissue damage.

In primary progressive (PP) or secondary progressive (SP) MS, phase II trials should also aim for a clinical outcome or change in an MRI metric of tissue damage, such as atrophy, T1 black hole volume, NAA spectroscopy or magnetization transfer imaging. Except for early SP MS where there may be concomitant frequent relapses, gadolinium enhancement would not be useful, nor would measures of relapse rate. One difficulty in trying to alter progressive disease lies in the clinical measurement of progression of disability. The commonest scale, the Kurtzke expanded disability status scale (EDSS) is rather insensitive, requires a minimum of six months to demonstrate a change, is weighted heavily toward ambulation and is not linear. A newer scale, the Multiple Sclerosis Functional Composite (MSFC) has several advantages over the EDSS, including linearity, need for fewer subjects to achieve the same power, and less variability, but, as opposed to the EDSS, has not yet been accepted as an outcome measure by regulatory authorities.

II.1 Outline of a typical development plan

A typical phase II trial testing an anti-inflammatory agent in RR MS will utilize a multicenter, randomised, double blind, placebo-controlled, parallel group design.

II.2. Short term studies

II.2.A. Objectives
To evaluate short term efficacy and safety of immunomodulatory or anti-inflammatory drug.

II.2.B. Primary Endpoints
Cumulative total number of enhancing lesions on all post Gd T1 weighted MRI images from monthly scans performed from week 0-12 to week 24-36. Other endpoints related to MRI disease activity can be selected as primary endpoints (see secondary endpoints).

II.2.C. Secondary Endpoints
MRI parameters:
   a. number of new enhancing lesions
   b. total volume of enhancing lesions
   c. number of new T2 weighted lesions
   d. number of T2 weighted lesions
   e. total volume of T2 weighted lesions
   f. number and volume of T1 weighted hypointense lesions
   g. other MRI parameters such as progression of brain atrophy, or variation of magnetisation transfer ratio parameters, (these last parameters particularly useful when testing a non anti-inflammatory drug).

Clinical parameters:
   a. Related to clinical relapses: number of relapses, relapse rate, time to first relapse, number of relapse-free patients
b. Related to disability progression: number of patients with a predefined interval EDSS progression (usually EDSS score ≥ 1), variations of MSFC scores.

Safety and tolerability parameters: incidence and prevalence.

II.2.D. Exploratory endpoints
Exploratory endpoints must be added according to the pharmacological properties of the drug.

II.2.E. Study design
Multicenter, randomised, double blind, placebo-controlled, parallel group design, is used, the number of arms depending on the different dosages which are evaluated. A screening phase is generally required, 4-12 weeks, with 1 or 2 MRI examinations for selecting MRI active patients. Enrichment of the population study with only MRI active patients enhances the power of the study and reduces the number of patients required. The duration of the study varies between 24 to 36 weeks, depending on the onset of monthly MRI evaluation. In some trials patients are evaluated monthly from baseline, but sometimes, when the full effect of the drug is delayed, patients are evaluated from 12th week to 36th week. A longer period of evaluation is not possible with a placebo arm. Such a design becomes more difficult to perform than a few years ago for ethical reasons and availability of MS patients.

Medication permitted is corticosteroid treatment for MS relapses: 1g methylprednisolone over 3 hours IV infusion/day for 3 or 5 days is usually the recommended treatment of severe relapses. All symptomatic treatments such as antispastic, anticholinergic, antidepressant, antiepileptic drugs and rehabilitation are usually permitted.

II.2.F. Study Population
Sufficient data are available to calculate accurately the number of patients required for the study according to the anticipated drug effect. As an example, 60 patients per group would be sufficient over 6 months to detect a 50% difference with power taken as 90% and a type one error =0.05 based on the hypothesis of 2.8 ± 3.7 new Gd enhancing T1 weighted lesions (see Sormani et al.).

II.2.G. Specific Inclusion Criteria
a. Patients who meet the diagnosis criteria for MS according to guidelines provided by McDonald et al.
b. Patients who present at least one T1 weighted Gd enhancing lesion on MRI performed in the screening period.
c. Patients with clinical disability measured by EDSS score between 0 and 5.5 inclusive.
d. Male and female MS patients aged between 18 – 55 years. Women must use an effective method of contraception if they are childbearing potential.

II.2.H. Specific Exclusion Criteria
a. Patients who present a progressive evolution course defined as a sustained progression of disability evaluated by EDSS score in the year preceding the screening period.
b. Patients with relapse in the 2 months period preceding baseline. This period may vary according to the drug.
c. Patients with previous treatment with immunosuppressive, immunomodulating or any investigational drug. According to this previous treatment and the drug tested, exclusion must be complete or a wash-out period of variable duration may be accepted.
d. Usual exclusion criteria such as patients with concomitant severe or unstable non neurological disease which would induce any risk for the patient.
e. Pregnant or breastfeeding women.
II.2.I. Specific criteria for early withdrawal and discontinuation
A list of withdrawal criteria is pre-established such as the following: consent withdrawn, severe progression of the disease requiring recommended treatments, serious adverse event other than a relapse, insufficient compliance to the treatment, inadequate concomitant therapy, occurrence of pregnancy.

II.2.J. Data analysis methods
For efficacy analysis the primary population is the intention to treat (ITT) population. The secondary populations is the per protocol population with no major protocol deviation. For safety analysis the exposed population is analysed. Multiple statistical methods are usually used for the primary and secondary end points. The tests are usually two-sided with a global type 1 error, $\alpha \leq 0.05$. These methods, adapted to the parameters analysed, must be predefined when the protocol is designed.

II.2.K. Extension studies
The goal of phase 2 studies are: confirmation of “a proof of concept” and/or safety requirements. In the majority of cases, these controlled trials are randomized against placebo, which forbids a long term study. There is no guideline to recommend for managing the patients at the end of the treatment period, the positions of ethical committees varying from country to country.

II.3. Phase 2 studies in primary progressive MS

PPMS remains the only subtype of MS for which there is no approved disease modifying therapy. Immunomodulators and immunossuppressive therapies are used in relapsing secondary progressive MS but these compounds are not effective in SP with sustained progression without relapses, which is similar to PPMS.

Tolerability and efficacy phase 2 studies are randomized, double blind, placebo controlled trials. MRI outcomes are not accepted as surrogate outcomes for phase 2 studies in PPMS Therefore the primary outcome generally used is the time to sustained treatment failure as defined by progression on disability scale, usually EDSS or a more sensitive scale, MSFC.

Some secondary MRI outcomes are available such as the measurement of progression of brain atrophy. The duration of the treatment period is 2 years. Methodology is as the whole similar to the phase 2 trials in RRMS.

III. PHASE III STUDIES FOR REGISTRATION OF NEW DISEASE MODIFYING DRUGS

III.1. Long-term studies to slow or halt relapsing-remitting MS

III.2. Outline of a typical developmental plan

A clinical developmental plan will include at least one large phase III study with a clinical primary efficacy outcome, either relapse rate or disease progression, with a study duration of at least 2 years. The pivotal trial for registration purposes has hitherto included one large well-conducted placebo-controlled trial. Initially, both the U.S. Food and Drug Administration (FDA) and the European Medicines Evaluation Agency (EMEA) have granted approval after one adequate and well controlled trial in patients with relapsing-remitting MS. However, for the latest approved disease-modifying drug in the United States, approval was granted after one large placebo-controlled trial and a short-term comparative trial with the brand leader. In relapsing-remitting multiple sclerosis placebo-controlled trials have become increasingly difficult to perform after approval of several disease modifying drugs for treatment of the disease activity. There are both ethical and practical issues involved. The ethical problems regarding
placebo-controlled trials in relapsing-remitting multiple sclerosis are in principal identical for phase II and phase III trials, whereas the practical problems with enrolment of large patient numbers for prolonged studies are by far more pronounced in phase III studies. Regarding ethical issues for future placebo-controlled clinical trials an international taskforce of clinicians, statisticians, ethicist and regulators concluded that placebo-controlled clinical trials in forms of multiple sclerosis for which partially effective therapies exist were ethical as long as study subjects were fully appraised of the availability of such therapies and were encouraged to pursue them outside of a clinical trial. Patients who declined to utilise available treatment after proper education and counselling, or those that failed all therapies can be considered to have no treatment alternatives and thus may participate in a placebo-controlled trial. Future requirements for approval of new disease modifying drugs in relapsing-remitting MS by the FDA or EMEA are not known. A developmental plan may include both superiority head to head trials and non-inferiority trials against an established approved drug. Regulatory authorities have not yet approved placebo-controlled trials with historical placebo-controls.

III.2.A. Objectives
To evaluate the efficacy and safety of the investigational drug as mono-therapy in patients with relapsing-remitting multiple sclerosis.

III.2.B. Primary Endpoints
The primary endpoint should be clinical. Time to progression in disability should be preferred in trials of 2-3 years duration. Progression is usually measured as increase of one full (1.0) step on Kurtzkes EDSS scale (0.5 step in patients with a baseline EDSS of 5.5 or above). Progression should be confirmed at 2 assessments with an interval of 3 or 6 months. Alternatively the multiple sclerosis functional composite (MSFC) be used, but this so far has no been accepted as a valid primary endpoint by regulatory authorities, mainly because it has not yet been defined how progression is measured on this scale, and how worsening on this scale should be explained clinically. Changes in the annual relapse rate can also be used as primary endpoint. Registration of confirmed relapses is preferred to the use of reported relapses with or without confirmation.

III.2.C. Secondary Endpoints
If progression is used as the primary endpoint, the annual relapse rate should be included as secondary endpoint, and vice versa. A surrogate marker, MRI, should always be included as secondary endpoint. Gadolium-enhancing lesions on T1-weighted images, new disease activity on T2-weighted images or combined unique activity (CUA), i.e. a combination of new or enlarging T2-lesions and gadolinium positive lesions, are recommended. Alternatively, changes in T2-lesion area or T1-lesion area (black holes) can be used. Recently, a brain atrophy measure e.g. the brain parenchymal fraction has gained use as a MRI secondary endpoint.

III.2.D. Exploratory Endpoints
Clinical exploratory endpoints are time to first relapse, proportion of relapse-free patients, integrated disability status score (IDSS), i.e. the area under the disability time curve, proportion of patients with progression, and time to EDSS 6 or 7. Exploratory MRI endpoints include some of the MRI endpoints measured under secondary endpoints, gadolinium enhancing lesions, and new magnetic resonance techniques like magnetisation transfer ratios and magnetic resonance spectroscopy.

III.2.E. Study design
A randomised double-blind, placebo-controlled parallel group design is used. The trial involves a baseline evaluation with control for fulfilling of inclusion and exclusion criteria. Patients should be assessed clinically with intervals of 3 months and in case of an acute relapse. Assessment includes scoring on the expanded disability status scale (EDSS) and/or the multiple sclerosis functional composite (MSFC) scale. MRI measures should be obtained at yearly intervals or at least at baseline and
study end. The duration of the study should be at least 2 years and preferably 3 years, if disease progression is the primary endpoint.

III.2.F. Planned sample
The sample size depends on the chosen minimal relevant deficit in the primary endpoint. Typically a sample size of above 300 patients per treatment-arm is required to detect 30-40% difference between the trial-arms with an 80% power and a type I error (two-sided) of 5%.

III.2.G. Study population
Patients with relapsing-remitting multiple sclerosis according to accepted criteria (McDonald criteria or Poser criteria) and age 18-55 years. Only patients with low or moderate disability (EDSS ≤ 5) should be included. Enrolled patients should have suffered recent disease activity, usually in the previous year, either clinical activity or MRI activity.

III.2.H. Specific inclusion criteria
Similar to those described for phase II placebo-controlled trial.

III.2.I. Specific exclusion criteria
Similar to those described for phase II placebo-controlled trial.

III.2.J. Specific criteria for early withdrawal and discontinuation
Placebo-controlled trials should include exit (escape) criteria defined as significance disease progression, e.g. 2 steps on EDSS, or frequent and severe relapse activity.

III.2.K. Data analysis methods
The analysis of the primary efficacy variable is based on the intention-to-treat (ITT) population. Time to progression is analysed using the log-rank test and Kaplan-Meier estimates.

IV. COMPARATIVE STUDIES

IV.1. Superiority studies
Criteria for conducting comparative superiority trials in relapsing-remitting multiple sclerosis have not been defined. There is currently no international gold standard for such trials. Below is an example that was accepted by the U.S. Food and Drug Administration as additional trial to a placebo-controlled study for approval of a disease-modifying agent. It has to be recognised that this study has been criticised in the medical and scientific community, mainly because its length was thought to be suboptimal.

IV.1.A. Objectives
To evaluate the comparative efficacy and tolerability of the investigational drug versus an approved active control in relapsing-remitting multiple sclerosis.

IV.1.B. Primary endpoints
Proportion of patients who remained relapse-free at 24 weeks.

IV.1.C. Secondary endpoints
Mean number of relapses per patient during 24 weeks, number of active lesions per patient per scan at 24 weeks on MRI.
IV.1.D. Exploratory endpoints
Mean number of combined unique activity lesions, i.e. a combination of new or enlarging T2-lesions and gadolinium positive lesions, per patient per scan, mean number of T1 active lesions per patient per scan, mean number of T2 active lesions per patient per scan, proportion of relapse-free patients at 48 weeks.

IV.1.E. Study design
A randomised parallel-group, single-blind study. Patients and treating physicians were aware of treatment allocation, whereas the evaluating neurologist and radiologist were blinded to study treatment. Ideally, the study should be double-blind but this might be difficult to achieve depending on the characteristics of the drugs under study. The primary efficacy endpoint was assessed at 24 weeks but the study was continued for 48 weeks.

IV.1.F. Planned sample
A sample size of 280 patients per treatment arm provided a 95% power at a significance level of 5% to detect a 30% relative increase in the primary endpoint.

IV.1.G. Study population
Patients with relapsing-remitting multiple sclerosis according to Poser criteria and age 18-55 years. Only patients with EDSS 0 to 5.5 and 2 relapses in the prior 2 years were included.

IV.1.H. Data analysis method
The analysis of the primary efficacy variable was based on the intention-to-treat (ITT) population. The primary end point, the odds ratio for remaining relapse free at 24 weeks, was analyzed by logistic regression with adjustment for treatment and centre.

IV.2. Non-inferiority studies

IV.2.A. Objectives
To evaluate the efficacy and tolerability of the investigational drug in patients with relapsing-remitting multiple sclerosis in comparison with an established drug at fully effective dosage under mono-therapy conditions.

IV.2.B. Primary endpoints
Primary endpoints for such studies have not been defined but should be a clinical measure assessed in the per-protocol (PP) population. Possible primary endpoints would include time to first relapse, proportion of relapse-free patients, annual relapse rate or time to progression on expanded disability status scale (EDSS).

IV.2.C. Secondary endpoints
Clinical endpoints include: Time to first relapse, relapse-free patients, annual relapse rate, time to progression on EDSS confirmed at 6 months, proportion of patients with progression.

MRI endpoints include: New disease activity on T2-weighted images or combined unique activity (CUA), e.g. a combination of new or enlarging T2-lesions and gadolinium positive lesions. Alternatively, changes in T2-lesion area, T1-lesion area (black holes), or brain atrophy can be used.

IV.2.D. Study design
The trial may be a multi-centre, double-blind, randomised parallel-group design with a double dummy technique comparing the investigational drug with the best reference treatment at optimised dosage. The double-blind phase may be followed by an open-labelled extension study.
IV.2.E. Planned sample
The authors are not aware of widely accepted sample size calculations for this type of study; it has to be
recognised, however, that by concept non-inferiority studies involve a huge number of patients.

IV.2.F. Study population
Patients with relapsing-remitting multiple sclerosis according to accepted criteria (McDonald criteria or
Poser criteria) and age 18-55 years. Only patients with low or moderate disability (EDSS ≤ 5) should be
included. Enrolled patients should have suffered recent disease activity, usually in the previous year, either
clinical activity or MRI activity.

IV.2.G. Specific criteria for early withdrawal or discontinuation
Placebo-controlled trials should include exit (escape) criteria defined as significance disease progression,
e.g. 2 steps on EDSS, or frequent and severe relapse activity.

IV.2.H. Data analysis method
In non-inferiority trials, analysis of the primary efficacy variable is made on the per-protocol (PP)
population. Relapse-free rates or progression-free rates may be compared by a logistic regression model
whose 95% confidence interval computation may include baseline characteristics as factors.

V. SECONDARY PROGRESSIVE MS

V.1. Outline of a typical developmental plan
The benefit of disease modifying therapies in patients with secondary progressive MS is less apparent.
The results in the European study with interferon-beta 1b showed a modest slowing of progression in
secondary progressive MS patients of whom many had relapses. By contrast, no effect on disability
progression was observed in the North American study of interferon-beta 1b or in the SPECTRIMS study
with interferon-beta 1a. In another study with interferon-beta 1a, the IMPACT study, only an effect on the
MSFC was found. Hence, it can be concluded that placebo-controlled double-blind trial are still ethical
and feasible in patients with secondary progressive multiple sclerosis. Patients with secondary progressive
MS who have still relapses should be informed about the possibility of starting with an approved drug
outside a clinical trial, and only patients who have declined to do so should be included in clinical trials.

V.2. Placebo-controlled trials

V.2.A. Objectives
To evaluate the efficacy and safety of the investigational drug as mono-therapy in patients with secondary
progressive multiple sclerosis.

V.2.B. Primary endpoints
The primary endpoint should be clinical. Time to progression in disability should be preferred in trials of
2-3 years duration. Progression is usually measured as increase of one full step on Kurtzkes EDSS scale
(0.5 step in patients with a baseline EDSS of 5.5 or above). Progression should be confirmed at 2
assessments with an interval of 3 or 6 months. In the future the multiple sclerosis functional composite
(MSFC) might provide a useful alternative, but so far this scale has not been accepted as primary outcome
measure by regulatory authorities.

V.2.C. Secondary endpoints
The annual relapse rate should be included as secondary endpoint. A surrogate marker, MRI, should
always be included as secondary endpoint. New disease activity on T2-weighted images or combined
unique activity (CUA), e.g. a combination of new or enlarging T2-lesions and gadolinium positive lesions
or brain atrophy e.g. measured as the brain parenchymal fraction, are recommended. Alternatively, changes in T2-lesion area or T1-lesion area (black holes) can be used.

**V.2.D. Exploratory endpoints**
Clinical exploratory endpoints are proportion of patients with progression, and time to EDSS 6 or 7, integrated disability status score (IDSS), e.g. the area under the disability time curve, time to first relapse, proportion of relapse-free patients. Exploratory MRI endpoints may include some of the MRI endpoints measured under secondary endpoints, gadolinium enhancing lesions, and new magnetic resonance techniques like magnetisation transfer ratios and magnetic resonance spectroscopy.

**V.2.E. Study design**
A randomised double-blind, placebo-controlled parallel group design is used. The trial involves a baseline evaluation with control for fulfilling of inclusion and exclusion criteria. Patients should be assessed clinically with intervals of 3 months and in case of an acute relapse. Assessment includes scoring on the expanded disability status scale (EDSS) and/or the multiple sclerosis functional composite (MSFC) scale. MRI measures should be obtained at yearly intervals or at least at baseline and study end. The duration of the study should be at least 2 years and preferably 3 years, if disease progression is the primary endpoint.

**V.2.F. Planned sample**
The sample size depends on the chosen minimal relevant deficit in the primary endpoint. Typically a sample size of above 300 patients per treatment-arm is required to detect 30-40% difference between the trial-arms with an 80% power and a type 1 error (two-sided) of 5%.

**V.2.G. Study population**
Patients with secondary progressive multiple sclerosis according to accepted criteria and age 18-55 years. Only patients with moderate disability (EDSS 3 to 5.5) should be included. Enrolled patients should have suffered recent clinical disease activity, i.e. progression or relapses during the last 1-2 years.

**V.2.H. Specific inclusion criteria**
Similar to those described for phase II placebo-controlled trial.

**V.2.I. Specific exclusion criteria**
Similar to those described for phase II placebo-controlled trial.

**V.2.J. Specific criteria for early withdrawal and discontinuation**
Placebo-controlled trials should include exit (escape) criteria defined as significance disease progression, e.g. 2 steps on EDSS.

**V.2.K. Data analysis methods**
The analysis of the primary efficacy variable is based on the intention-to-treat (ITT) population. Time to progression is analysed using the log-rank test and Kaplan-Meier estimates.

**VI. PRIMARY PROGRESSIVE MS**

**VI.1. Outline of a typical developmental plan**
The clinical course in primary progressive MS is characterized by a progressive accumulation of neurological deficits from onset without relapses. There is no approved therapy for this course of multiple sclerosis.

A clinical developmental plan will include at least one large phase III study with disease progression as the clinical primary efficacy outcome, and with a study duration of at least 2 years.
VI.2. Placebo-controlled trials

VI.2.A. Objectives
To evaluate the efficacy and safety of the investigational drug as mono-therapy in patients with secondary progressive multiple sclerosis.

VI.2.B. Primary endpoints
The primary endpoint should be clinical, and time to progression in disability should be preferred in trials of 2-3 years duration. Progression is usually measured as increase of one full step on Kurtzke’s EDSS scale (0.5 step in patients with a baseline EDSS of 5.5 or above). Progression should be confirmed at 2 assessments with an interval of 3 or 6 months. Alternatively, the multiple sclerosis functional composite (MSFC) may be used, but it has not yet been approved as primary outcome measure.

VI.2.C. Secondary endpoints
A surrogate marker, MRI, should always be included as secondary endpoint. The currently recommended measures for therapeutic trials in relapsing remitting and secondary progressive multiple sclerosis show only little change in primary progressive multiple sclerosis and therefore more pathologically specific MRI measures may be required. Changes in T1-lesion area (black holes), brain atrophy, e.g., measured as the brain parenchymal fraction or cervical cord cross-sectional area may be used.

VI.2.D. Exploratory endpoints
Clinical exploratory endpoints are proportion of patients with progression, and time to EDSS 6 or 7. Exploratory MRI endpoints may include some of the new magnetic resonance techniques like magnetisation transfer ratios and magnetic resonance spectroscopy.

VI.2.E. Study design
A randomised double-blind, placebo-controlled parallel group design is used. The trial involves a baseline evaluation with control for fulfilling of inclusion and exclusion criteria. Patients should be assessed clinically with intervals of 3 months and in case of an acute relapse. Assessment includes scoring on the expanded disability status scale (EDSS) and/or the multiple sclerosis functional composite (MSFC) scale. MRI measures should be obtained at yearly intervals or at least at baseline and study end. The duration of the study should be at least 2 years and preferably 3 years, if disease progression is the primary endpoint.

VI.2.F. Planned sample
The sample size depends on the chosen minimal relevant deficit in the primary endpoint. It has been reported that in secondary progressive multiple sclerosis, a sample size of above 300 patients per treatment-arm is required to detect 30-40% difference between the trial-arms with an 80% power and a type 1 error (two-sided) of 5%. However, such information is currently not available for primary progressive multiple sclerosis.

VI.2.G. Study population
Patients with primary progressive multiple sclerosis according to accepted criteria and age 18-55 years should be studied. It is a problem that patients with primary progressive multiple sclerosis do not readily conform to accepted criteria (McDonald criteria or Poser criteria) and have a wide differential diagnosis. Only patients with moderate disability (EDSS 3 to 5.5) should be included. Enrolled patients should have suffered recent clinical disease activity, i.e., progression or relapses during the last 1-2 years.

VI.2.H. Specific inclusion criteria
Similar to those described for phase II placebo-controlled trial.
VI.2.I. Specific exclusion criteria
Similar to those described for phase II placebo-controlled trial.

VI.2.J. Data analysis methods
The analysis of the primary efficacy variable is based on the intention-to-treat (ITT) population. Time to progression is analysed using the log-rank test and Kaplan-Meier estimates.

VII. OTHER STUDIES (ATYPICAL MS FORMS)
Rare inflammatory demyelinating syndromes, such as Marburg’s Disease, Pseudotumoral forms, Baló’s Concentric Sclerosis (BCS) or Devic’s Neuromyelitis Optica (DNO), are difficult to be studied in a randomised controlled clinical trial mainly due both to absence of specific clinical and laboratory findings leading to an early and definitive diagnosis and to the small number of patients seen every year with homogeneous clinical characteristics (for example at the onset of disease, after a few clinical relapses and low disability). As a consequence, to date no reliable data are available on the efficacy of immunomodulatory or immunosuppressive drugs, usually used in MS, in halting or slowing the inflammatory and degenerative processes underlying these rare inflammatory diseases.

VII.1. Marburg’s disease
Marburg’s disease is an acute malignant monophasic demyelinating disease, which usually results in death within a few weeks or months after the initial bout. It is characterized by widespread and progressive cerebral white matter destruction or by severe pathological involvement of clinically strategic regions such as brainstem, resulting in bulbar paralysis. Short-term, observational studies of a restricted cohort of patients evaluating the therapeutic efficacy of the association of high dosage Metilprednisolone and frequent, pulse administration of immunosuppressive drugs (cyclophosphamide or mitoxantrone) or plasmapheresis (with or without subsequent pulse cyclophosphamide) have been reported.

VII.2. Baló’s Concentric Sclerosis and Pseudotumoral forms
Baló’s Concentric Sclerosis is a rare demyelinating disorder characterized pathologically by concentric rings of alternating demyelinated and relatively myelin preserved white matter. The pathogenesis of the concentric lesion may be explained by periodic suppression of demyelination in a rapidly expanding area of inflammation, allowing remyelination or only transient incomplete demyelination to occur. While initial reports of BCS predicted that the disease was rapidly progressive and fatal, a good prognosis has been recently described in a few cases and the lesion is usually considered as an atypical, large, pseudotumoral MS plaque. Large, focal, tumor-like demyelinating lesions of the CNS often represent a diagnostic challenge, which reasonably calls for a stereotactic biopsy, particularly when isolate in the brain, to exclude glioma, infectious processes or primary CNS vasculitis. These cases are usually characterized by a severe course and a rapid clinical deterioration. Response to therapy is highly variable. Patients are usually treated with steroid, generally in high doses intravenously, as well as variably with cyclophosphamide or plasma exchange. Despite the aggressive therapy sometimes the disease ultimately progresses, mainly in patients with spinal cord involvement.

VII.3. Devic’s Neuromyelitis Optica (DNO)
DNO is usually considered as a distinct disease entity from Ms according to several clinical, neuroradiological and CSF findings. Pathological aspects and the relevant pathogenetic role of humoral immunity also support the belief that DNO may be a separate syndrome. DNO diagnosis is difficult to make after the first episode. Two or more acute episodes of neurological impairment involving the optic nerves and the spinal cord, in a simultaneous or sequential temporal relationship, must be observed, with no clinical or MRI evidence of brain involvement. A poor prognosis may be predicted by the age at onset, the interval between first and second episode, the relapse rate and the severity of the first attack. At the moment no effective treatment has been demonstrated, although chronic or pulse steroid therapy is often used in DNO
patients, associated with a variety of immunosuppressive or immunomodulatory agents, empirically chosen, usually after 3-5 attacks have occurred.

**VII.4. Future therapeutic strategies**

New emerging disease-modifying therapies that target cytokines, the blood brain barrier, the “trimolecular complex” or that act by deletion of auto-reactive T cells are candidates for treatment of these rare MS variants and their efficacy and safety should also be evaluated. These syndromes are characterised by a very aggressive course, therefore the research of active and efficacious treatments needs to be strongly encouraged. Since most of these patients have usually a bad prognosis and deteriorate to severe irreversible disability in only a few months or years, beneficial effect due to therapeutic intervention may be easily detected by using strong primary end points, such as death or loss of deambulation. Classical phase II-III studies are impossible in this MS variants; small group of patients treated and monitored with a specific protocol are the only feasible approach, even if it implies problems with interpretation and generalization of the results.

The Restricted Cohort Study applies the principles and patient enrolment procedures regularly used in randomised clinical trials. Therefore, strict eligibility criteria and the appropriate choice of zero time must be well defined; anyway baseline differences should be adjusted for prognostic risk. In addition, patients must be classified according to suitable clinical criteria to enable adjustment for any inequalities in susceptibility to the outcome. Finally, the analysis of the data should be conducted using the same methodology used in clinical trials.

The primary objective of these studies is to provide preliminary data on the efficacy, considered both on short term evolution of the recent clinical attack as well as on long term evolution of the disease, evaluated as confirmed changes of EDSS and ambulation index. MRI and neurophysiological test are very important to support clinical observation. The secondary objectives of these studies are to gather descriptive information concerning short and long term tolerability and safety of the investigational therapeutic intervention as well as about the potential loss of efficacy on long-term chronic or pulse administration.

**VIII. EXAMPLES OF PHASE III TRIALS IN MULTIPLE SCLEROSIS**

**Pivotal placebo-controlled trials in relapsing remitting multiple sclerosis**


6. Comi G, Filippi M, Wolinsky JS. European/Canadian multicenter, double-blind, randomized, placebo-controlled study of the effects of glatiramer acetate on magnetic resonance imaging--measured disease

**Comparative superiority Phase III trials in relapsing remitting multiple sclerosis**


**Phase III placebo-controlled trial in early multiple sclerosis**


**Phase III placebo-controlled trials in secondary multiple sclerosis**


**IX. SUGGESTED READINGS**

Pharmacological Research in Mental Disorders

Chapter 21. Mood Disorders

Chapter 22. Anxiety Disorders

Chapter 23. Schizophrenic Disorders

Chapter 24. Alcoholism and Nicotine Addiction

**GLOSSARY OF TERMS**

<table>
<thead>
<tr>
<th>Term</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>ADS</td>
<td>Alcohol Dependence Scale</td>
</tr>
<tr>
<td>AIMS</td>
<td>Abnormal Voluntary Movement Scale</td>
</tr>
<tr>
<td>ARCI</td>
<td>Addiction Research Center Inventory</td>
</tr>
<tr>
<td>BAS</td>
<td>Barnes Akathisia Scale</td>
</tr>
<tr>
<td>BDI</td>
<td>Bipolar I Disorder</td>
</tr>
<tr>
<td>BDI</td>
<td>Beck Depression Inventory</td>
</tr>
<tr>
<td>BSRS</td>
<td>Brief Psychiatric Scale</td>
</tr>
<tr>
<td>CGI</td>
<td>Clinical Global Impression Scale</td>
</tr>
<tr>
<td>CIWA</td>
<td>Clinical Institute Withdrawal Assessment for Alcohol</td>
</tr>
<tr>
<td>DRUG AAA</td>
<td>Investigational Drug for Treatment of Alcohol Dependence</td>
</tr>
<tr>
<td>DRUG SSS</td>
<td>Investigational Drug for Treatment of Nicotine Dependence</td>
</tr>
<tr>
<td>DRUG XOXO</td>
<td>Investigational Drug for Treatment of MDD</td>
</tr>
<tr>
<td>DRUG XXX</td>
<td>Investigational Drug for Treatment of Anxiety</td>
</tr>
<tr>
<td>DRUG YYY</td>
<td>Investigational Drug for Treatment of Schizophrenia</td>
</tr>
<tr>
<td>DSM-IV</td>
<td>Standard Diagnostic Criteria for Mental Disorders</td>
</tr>
<tr>
<td>ESRS</td>
<td>Extrapyramidal Symptom Rating Scale</td>
</tr>
<tr>
<td>FTND</td>
<td>Fagerstrom Test for Nicotine Dependence</td>
</tr>
<tr>
<td>GAD</td>
<td>General Anxiety Disorder</td>
</tr>
<tr>
<td>HAM-A</td>
<td>Hamilton Rating Scale for Anxiety</td>
</tr>
<tr>
<td>HAM-D</td>
<td>Hamilton Rating Scale for Depression</td>
</tr>
<tr>
<td>MADRS</td>
<td>Montgomery Asberg Depression Scale</td>
</tr>
<tr>
<td>MAOI</td>
<td>Monoamine Oxidase Inhibitors</td>
</tr>
<tr>
<td>MAST</td>
<td>Michigan Alcohol Screening Test</td>
</tr>
<tr>
<td>MDD</td>
<td>Major Depressive Disorder</td>
</tr>
<tr>
<td>OBS</td>
<td>Obsessive Compulsive Syndrome</td>
</tr>
<tr>
<td>PANSS</td>
<td>Positive and Negative Symptom Scale</td>
</tr>
<tr>
<td>PD</td>
<td>Panic Disorder</td>
</tr>
<tr>
<td>POMS</td>
<td>Profile of Mood States</td>
</tr>
<tr>
<td>SAS</td>
<td>Simpson Angus Scale</td>
</tr>
<tr>
<td>SHAPS</td>
<td>Snaith-Hamilton Pleasure Scale</td>
</tr>
<tr>
<td>SNRI</td>
<td>Serotonin-Norepinephrine Re-uptake Inhibitors</td>
</tr>
<tr>
<td>SSRI</td>
<td>Selective Serotonin Re-uptake Inhibitors</td>
</tr>
<tr>
<td>STA-IX</td>
<td>Stait-Trait Anxiety Scale</td>
</tr>
<tr>
<td>TCA</td>
<td>Tricyclic Antidepressants</td>
</tr>
<tr>
<td>VAAS</td>
<td>Visual Analogue Anxiety Scale</td>
</tr>
<tr>
<td>VAS</td>
<td>Visual Analogue Scale</td>
</tr>
</tbody>
</table>
Chapter 21. Mood Disorders

Claudio A. Naranjo, M.D. ¹

Lara Chayab ²

¹ Professor
Department of Pharmacology
Psychiatry and Medicine
University of Toronto
Toronto, Ontario
CANADA
and
Head
Neuropsychopharmacology Research Program
Sunnybrook & Women’s College Health Science Centre
Toronto, Ontario
CANADA

² M.Sc. Candidate
Department of Pharmacology
University of Toronto
Toronto, Ontario
CANADA
and
Researcher
Neuropsychopharmacology Research Program
Sunnybrook & Women’s College Health Science Centre
Toronto, Ontario
CANADA
1. INTRODUCTORY REMARKS

The mood disorders that have been most extensively examined in epidemiological studies are major depressive disorder (MDD), dysthymia, and bipolar I disorder (BDI) (1). MDD is the most prevalent of the other disorders. The lifetime prevalence rates of MDD show that 5%-12% of men will experience depression at some point in their lifetime while that rate is higher in women accounting for 10%-25% (2). The diagnostic criteria of MDD according to the DSM-IV (American Psychiatric Association, 1994) are the feeling of sadness and/or loss of pleasure present most of the day, everyday, for at least two weeks (anhedonia). During this period, at least five other symptoms must be present including appetite disturbances, weight disturbances, sleep disturbances, activity disturbances, fatigue, inappropriate guilt, and thoughts of death (3). The lifetime prevalence rate of Bipolar I Disorder is much less than the prevalence of MDD accounting for only 1 to 2% (2). However, individuals exhibiting BDI experience more depressive episodes than those with MDD do (2). The diagnostic criteria for BDI according to the DSM-IV are the occurrence of one or more manic episodes or mixed episodes. Often individuals have also had one or more Major Depressive Episodes. Pharmacotherapy for BDI includes mood stabilizers, such as lithium and valproate, as first line of treatment and mood stabilizers with other medications for people unresponsive to first line treatments (1). During a manic episode, individuals experience hyperactivity, hallucinations, and paranoia (2). These symptoms usually cause problems to the diagnosed patients with the law, at work, and with other individuals. Because of the nature of the disorder, clinical studies aimed at investigating treatments for BDI are difficult to establish because of the lack of compliance and the large dropout rate.

Choices of pharmacological therapy for the treatment of depression include tricyclic antidepressants (TCAs), monoamine oxidase inhibitors (MAOIs), and newer agents such as selective serotonin reuptake inhibitors (SSRIs), and serotonin-norepinephrine reuptake inhibitors (SNRI) (1). TCAs and MOAIs have been associated with severe adverse effects, drug interactions, and toxicities. Side effects of TCAs administration include dry mouth, constipation, blurred vision, sedation, weight gain, and sexual dysfunction while the most frequent adverse effects of MAOIs are similar to TCAs and also include orthostatic hypertension, palpitations, tachycardia, peripheral edema, and muscle cramps (2). Depressed patients usually terminate the drug therapy due to the side effects before the full course of the treatment is achieved, which leads to recurrence of symptoms. SSRIs and SNRIs have a better profile of adverse effects. They are less sedating, and have no cardiac effects. However, they still cause nausea, headaches, insomnia, sexual dysfunction, tremors, and CYP 450 inhibition (2). If lower doses of the drugs are used to minimize adverse effects, the efficacy of the drugs will be negatively affected and full recovery will not be achieved. In addition, approximately 5 weeks are required for the onset of action of SSRIs, and sometimes about 1 week for the onset of action of SNRIs (2). As a result, there is a need for newer antidepressant agents that have great efficacy for moderate to severe depression, better profile of adverse effects, fewer drug interactions, and a faster onset of action.

In this chapter, a double-blind, placebo-controlled, parallel-group study design will be used to test the efficacy and safety of investigational drug XOXO for patients with moderate to severe depression. Previous clinical studies have shown that drug XOXO is a highly selective serotonin reuptake inhibitor, is linear and dose-proportional, has a half-life of 27-32 hrs which accounts for its once daily dosing, its onset of action occurs within one week, and is eliminated by biotransformation. In addition, drug XOXO has a low potential for drug-drug interactions since its effect on CYP 450 is negligible, and has a favorable profile of adverse effects. In this study design, these properties of the drug will be better characterized and its efficacy will be determined. The example of the study design illustrated in this chapter is most applicable to trials for evaluating antidepressants intended to be used in patients as first line of treatment.
II. PHASE II STUDIES FOR REGISTRATION OF NEW ANTIDEPRESSANT DRUGS

II.1. Outline of a typical development plan

This study will examine the efficacy and safety of drug XOXO in men and women experiencing moderate to severe depression. All patients enrolled in the study meeting inclusion/exclusion criteria and that give consent will be randomly assigned to receive one oral daily dose of drug XOXO or placebo for 12 weeks. Efficacy and safety measures are going to be performed at weekly intervals up to the end of week 12. The study will be a randomized, double blind, placebo-controlled and patients will come back for two follow-up assessments.

II.2. Short-term studies

II.2.a. Study Objectives

Primary Objectives
a. To compare the efficacy of drug XOXO treatment versus placebo in reducing the symptoms of MDD
b. To compare the safety of drug XOXO treatment versus placebo

Secondary Objectives
a. To determine the onset of the antidepressant action of drug XOXO
b. To determine the duration of the antidepressant effect of drug XOXO
c. To determine the peak antidepressant effects of drug XOXO

II.2.b. Primary Endpoints

The efficacy of antidepressant drugs in clinical trials is measured using a wide variety of assessment tools, which include clinical observations, interviews, and self-reports. Currently, a number of rating scales exists that provide a standardized approach to evaluate the severity of mental disorders and the treatment outcomes. Scales are designed to measure either general symptoms or disease-specific symptoms. Some scales have to be rated by psychiatrists; nurses or research assistants can rate others, and yet other scales are self-evaluated. The choice of the appropriate scales for the diagnosis of specific mental disorder and the evaluation of the efficacy of investigational drugs depends on the specificity, sensitivity, and simplicity of the scales in question. The assessment tools have been extensively researched and evaluated for their specificity and sensitivity for each of the mental disorders that are going to be discussed in this chapter.

Rating scales will be administered to assess the following dependent variables:

a. Structured Clinical Interview for DSM-IV to diagnose patients with major depressive disorder
b. The Montgomery Asberg Depression Scale, MADRS, (Score \( \geq 30 \) for severely depressed patients) and The Beck Depression Inventory, BDI, (Score \( \geq 16 \) for severely depressed patients) to assess current level of depression (4).
c. The Hamilton Depression Scale, HAM-D to assess severity of depressed mood (score \( \geq 17 \) for severely depressed patients), which contains 17 items to assess depressed mood, suicidal ideation, somatic symptoms, and loss of interest. Four additional items are included (i.e. diurnal variation, derealization, paranoid symptoms and obsessional symptoms), making the total questionnaire 21 questions in length (4).
d. The Beck Anxiety Inventory and/ Stait-Trait Anxiety, STA-IX, to exclude patients with comorbid depression and anxiety, which is composed of 21 questions and evaluates the current level of depression (4).
Responders to the antidepressant drug XOXO versus placebo, where a response is defined as:

a. $R = a$ reduction from baseline (weeks 1) on weeks 2, 4, 6, 8, 10, 12 during and post treatment as measured by the HAM-D.

b. $R = a$ reduction from baseline (week 1) on weeks 2, 4, 6, 8, 10, 12 during and post treatment as measured by the BDI.

c. $R < 2$ on weeks 2,4,6,8,10,12 during and post treatment, as measured by the Snaith-Hamilton Pleasure Scale (SHAPS), which is a validated self-assessment scale estimating the degree to which a person is able to experience pleasure or the anticipation of a pleasurable event (i.e. hedonic tone). A score of 2 or more "disagree/definitely disagree" is considered to be indicative of an anhedonic state (4).

d. $R = a$ reduction in the score on weeks 2,4,6,8,10,12 during and post treatment, as measured by the Addiction Research Center Inventory (ARCI), which is a 77-item questionnaire that measures subjective effects of drugs’ positive/reinforcing (e.g. euphoria, stimulation) and negative or dysphoric (e.g. sedation, confusion). This inventory allows the quantification of subjective drug effects with scales sensitive to the effects of specific drugs and drug classes (4).

e. $R = a$ reduction in the score on weeks 2,4,6,8,10,12 during and post treatment, as measured by the Profile of Mood States (POMS), which is commonly used for assessing drug-induced changes in mood, the POMS consists of a series of 72 adjectives. With respect to each adjective, subjects respond how they feel using a five-point scale ranging from "extremely" to "not at all". Tension-Anxiety, Anger-Hostility, Depression-Depression, Friendliness, Fatigue, Confusion, Vigor, Elation, Arousal, and Positive Mood are the 10 scales covered in the POMS (4).

II.2.c. Secondary Endpoints

a. The time of onset of a consistent decrease in depressed mood as measured by the Visual Analogue Scale (VAS) compared to baseline. The VAS is often used in the assessment of momentary changes in affect. They consist of a selection of visual analog rating scales (100mm lines) anchored at each end by opposing adjectives to evaluate drug "liking", drug effect and desire to experience the drug effects again. Subjects are instructed to rate how they feel by making a mark anywhere along the line (4).

b. The treatment day during which the greatest reduction in mood is present as measured by the HAM-D.

c. The number of patients that achieve HAM-D \leq 7.

d. Adverse effects are measured by changes in blood pressure, heart rate, GI motility, and abnormal laboratory tests (blood and liver).

II.2.d. Study Design

This is a double blind, randomized, placebo-controlled, parallel-group study. There will be 2 groups in this study, a moderate to severely depressed group, and a healthy control group. Patients in each group will receive either a single dose of drug XOXO or a single dose of placebo randomly once daily for 12 weeks. The number of patients receiving drug XOXO will equal the number of patients receiving placebo within each group. The antidepressant effects of drug XOXO will be measured using various questionnaires (discussed below). Objective measures, such as blood pressure, will also be obtained. The results will be compared and analyzed between and within groups.

Screening for Eligibility (Visit 1, Week 1)

Subjects will be interviewed to ensure suitability for study participation a week before the commencement of the study. The following will be required to assess eligibility:

a. Written informed consent.

b. Structured Clinical Interview for DSM-IV to assess psychiatric status and to rule out dependence on psychoactive substances.
c. Current level of depression (Hamilton Depression Scale (HAM-D), Montgomery Asberg Depression Scale (MADRS), and Beck Depression Inventory (BDI)), and anxiety (Beck Anxiety Inventory and/ State-Trait Anxiety (STA-IX)). This is the baseline measure to which all upcoming results will be compared against.

d. Brief medical examination (heart rate, blood pressure).

e. Medical history.

f. Review of inclusion/exclusion criteria.

g. Pregnancy test for women.

h. Blood and urine samples to assess liver function, hematology, biochemistry, and to detect the presence of other psychoactive drugs.

i. Participants will be provided with a pager number to be used if they experience any serious side effects and a wallet card containing information about participation in this study.

j. Patients are also given a diary in which they record drug compliance and any side effects that they may experience on a daily basis.

Treatment Phase (Visits 2 – 7, Weeks 2-12)

a. Eligible subjects will attend six treatment sessions (one every two weeks).

b. Medical examination.

c. Medication will be dispensed (enough pills for two weeks).

d. Treatment will take place at 2-week intervals consisting of 30 to 45 minute sessions with the research assistant.

e. A psychiatrist will be available for consultation, assessment, and treatment as needed (i.e. adverse drug reaction, increases in severity of depressive symptomatology).

f. Review Daily Diary forms on which patients record compliance with medication.

g. At each visit, the MADRS, BDI, HAM-D, SHAPS, ARCI, POMS, and VAS will be completed and subjects will be interviewed regarding concomitant illness and medication use.

h. Ask patients to return any unused medication in the vial.

i. Blood will be drawn for trough drug concentrations at visits 3, 5 and 7 (4, 8 and 12 weeks after commencing medication).

j. Blood and urine will also be collected at visits 4, 6 and 8 for drug screen, complete biochemistry and hematology analysis.

k. Subjects will be referred to their family physicians either at the end of the 12 week study or if a subject decides to terminate participation in the study.

l. Individuals who do not respond to drug XOXO will be referred to alternate psychiatric treatment or to their family physicians.

Follow-up visits (Visits 8-9, at 3 months and 6 months after treatment)

a. Review daily diary

b. Medical examination

c. Psychiatrist: examine any increased depressed symptoms, interview patients for concomitant illness and examine potential adverse reactions.

d. Complete questionnaires: MADRS, BDI, STAI-XI, HAM-D, SHAPS, ARCI, POMS, and VAS.

e. Blood and urine collection for drug screen, complete biochemistry, and hematoloy analyses.

II.2.e. Planned Sample

Flemming’s Single Stage Procedure will be used in calculating the sample size in the demonstrated phase II study (5). The procedure depends on the assumption that investigators usually have some knowledge of the activity of drugs similar to the one being studied. Therefore, in this study, researchers will compare the anticipated response of drug XOXO to other observed responses of similar drugs with the same
therapeutic indication. Researchers will then specify a probability of a response, which could then be compared to the actual responses to standard treatments. If the response exceeds that of standard treatments, then it can be concluded that Drug XOXO exhibits efficacy (5).

Therefore, assume:
Largest response proportion = Ro
Smallest response proportion = Ra
Hypothesis:
\[ R \leq Ro \]
\[ R \geq Ra \]
\[ \alpha, \text{ probability of rejecting the hypothesis of } R \leq Ro \]
\[ \beta, \text{ probability of rejecting the hypothesis of } R \geq Ra \]
For N patients recruited for phase II trial, the observed number of patient responses r has a binomial distribution with parameter \( \pi \).
Therefore, the sample size required for Flemming’s Single Stage Procedure is approximately,
\[ N = \frac{Z_{1-\alpha} \sqrt{Ro(1-Ro)} + Z_{1-\beta} \sqrt{Ra(1-Ra)}}{Ra-Ro}^2 \]

A treatment regimen using SSRIs indicated in the treatment of MDD in phase II studies is expected to yield a response in at least 35% of the patients being tested to show efficacy. Previous phase I trials have shown that Drug XOXO exhibits a higher safety profile than standard treatments indicated for MDD. A one-sided test size is set at 5% and the power at 80%. Since the new investigational drug is shown to be safer than the standard treatments, the values of Ro and Ra will be set at 0.6 and 0.5 respectively with \( \alpha = 0.05 \) and \( 1-\beta = 0.8 \). Therefore, from Table 12.1 in the “Statistical Tables for the Design of Clinical trials” handbook, or from calculating the equation, the sample size N will equal 16 (5). Therefore, at least 16 patients are needed in each group (16 in the MDD group and 16 in the healthy control group) to detect significance in efficacy for this trial. Therefore, a total of 38 patients are going to be enrolled for the successful completion of this study.

II.2.f. Study Population
Male or Female over 18 years of age meeting DSM-IV criteria for MDD and who exhibit moderate to severe depression or a Ham-D score of \( \geq 17 \).

II.2.g. Specific Inclusion Criteria
A subject will be eligible for inclusion in the study only if all of the following criteria apply:

a. Males or females between 19 to 50 years of age.
b. Socially stable.
c. Meet DSM-IV criteria for major depressive disorder.d. In-patients or out-patients.e. Non-smokers.

II.2.h. Specific Exclusion Criteria
Exclusion criteria must take into account the characteristics of the drug (pharmacokinetics, pharmacodynamics, drug-drug interactions, and adverse effects). Patients will not be eligible to participate in the study if one of the following criteria apply:

a. If meet criteria for Bipolar Disorder, schizophrenia, schizo-affective or other substance abuse/dependence.
b. Evidence of medical or surgical illness requiring treatment.c. History of psychoactive drug dependence or a positive urine test for psychoactive drugs.d. Use of medications which may interfere with the study procedures (e.g. SSRIs).
e. Any clinically significant abnormality evident in biochemistry or hematology test results or in urine analysis requiring further investigation.

f. Active suicidal ideation.

g. Receiving or will receive other investigational drug during the study.

h. Pregnant or lactating females.

II.2.i. Tools to assess endpoints

Efficacy should be evaluated using the tools depicted in the following table:

<table>
<thead>
<tr>
<th>Variable</th>
<th>Assessment Tools to Measure Variable</th>
<th>Time of Assessment</th>
</tr>
</thead>
<tbody>
<tr>
<td>MDD diagnosis</td>
<td>DSM-IV, HAM-D</td>
<td>Visit 1, Week 1</td>
</tr>
<tr>
<td>Level and severity of depression</td>
<td>HAM-D, MADRS, BDI</td>
<td>Visit 1, Week 1</td>
</tr>
<tr>
<td>Excluding patients with concomitant anxiety</td>
<td>STA-XI</td>
<td>Visit 1, Week 1</td>
</tr>
<tr>
<td>Reduction in depressed mood</td>
<td>HAM-D, POMS</td>
<td>Visits 2 - 9</td>
</tr>
<tr>
<td>Reduction in loss of interest (ability to experience pleasure)</td>
<td>HAM-D, SHAPS</td>
<td>Visits 2 - 9</td>
</tr>
<tr>
<td>Increase in Euphoria</td>
<td>HAM-D, ARCI</td>
<td>Visits 2 - 9</td>
</tr>
<tr>
<td>Decrease in dysphoria</td>
<td>ARCI</td>
<td>Visits 2 - 9</td>
</tr>
<tr>
<td>Increase/decrease in hostility, fatigue, and confusion</td>
<td>POMS</td>
<td>Visits 2 - 9</td>
</tr>
<tr>
<td>Time of onset of consistent decrease in depressed mood</td>
<td>VAS, HAM-D</td>
<td>Visits 2 - 9</td>
</tr>
<tr>
<td>Time of greatest reduction in depressed mood</td>
<td>HAM-D</td>
<td>Visits 2 - 9</td>
</tr>
<tr>
<td>Number of patients achieving HAM-D ≤ 7 after end of study</td>
<td>HAM-D</td>
<td>Visits 2 - 9</td>
</tr>
<tr>
<td>Drug Compliance</td>
<td>Daily Diary and returned pills</td>
<td>Visits 2 - 7</td>
</tr>
</tbody>
</table>

Tools to assess safety

Adverse events such as GI abnormalities, blood pressure and heart rate changes, and blood biochemistry and hematology changes will be assessed in this study.

a. A complete medical examination will be performed, by a physician, at baseline (visit 1) as well as during all visit days. Any changes in blood pressure, heart rate, GI motility, and other complaints made by the patient will be recorded and compared to baseline. In the case of a patient developing any kind of adverse reaction, the subject will be immediately asked to return all medications and withdraw from the study. The patient will then be referred to his/her family doctor to avoid further complications.

b. Urine and blood tests will be performed on visits 2, 4, 6, 8, and 9. Any changes will also be reported and the patient will be asked to withdraw from the study.

c. The daily diaries are provided for the patients to record their feelings, drug compliance, and the occurrence of any adverse effects daily. The diaries will then be reviewed by the psychiatrist and their contents discussed by the patients.

d. If a serious side effect develops in a patient, a full analysis will be made to ensure that the adverse effect is from the investigational drug and not caused by other drugs that the patient may have taken, drug interaction, or a disease.

e. The adverse effects caused by the investigational drug XOXO and those developing from placebo will be compared to determine if a significant difference exists in order to identify the safety profile of drug XOXO.
II.2.j. Specific criteria for early withdrawal and discontinuation

Subjects are allowed to withdraw from the study at any time. Subjects must leave the study if one of these conditions holds:

a. Occurrence of serious side effects
b. Pregnancy
c. Non-compliance
d. Development of a medical condition
e. Use of other medication
f. Violation of the protocol
g. Withdrawal of consent

Subjects that terminate their involvement in the study because of the occurrence of side effects will be considered as having completed the study. Blood tests, urine tests and a complete medical exam should be performed on these patients. The results should determine whether the patients need to be placed on therapy to eliminate the side effects or whether the side effects will resolve on their own. In addition, these patients will be referred to their family physician and an assessment session should be performed after 3 months of withdrawal. Subjects that withdraw because of use of other medications, non-compliance, violation of the protocol, pregnancy, or development of a medical condition will be considered as having not completed the study and will be replaced by new participants. Subjects that terminate the study after drug administration should be contacted and followed-up to ensure that no severe side effects or worsening of the condition takes place. These subjects should also be referred to their family physician and called for a follow-up assessment after 3 months.

II.2.k. Data Analysis Method

For each visit, subject group (ex. Depressed vs. Controls) and dependent variables (e.g. HAM-D scores), parameters such as the mean, maximum, minimum change from baseline will be calculated and analyzed for the effects of drug XOXO. An analysis of variance will be conducted to compare all groups in order to determine if there is a significant difference in the way the groups responded to drug XOXO challenge. These data will also be entered into an ANOVA in order to evaluate the role of depression on the effect of drug XOXO. Together, these analyses should provide information on the efficacy and safety of drug XOXO on patients with MDD.

III. REFERENCES

Chapter 22. Anxiety Disorders

Claudio A. Naranjo, M.D. ¹

Lara Chayab ²

¹ Professor
Department of Pharmacology
Psychiatry and Medicine
University of Toronto
Toronto, Ontario
CANADA
and
Head
Neuropsychopharmacology Research Program
Sunnybrook & Women’s College Health Science Centre
Toronto, Ontario
CANADA

² M.Sc. Candidate
Department of Pharmacology
University of Toronto
Toronto, Ontario
CANADA
and
Researcher
Neuropsychopharmacology Research Program
Sunnybrook & Women’s College Health Science Centre
Toronto, Ontario
CANADA
I. INTRODUCTORY REMARKS

Anxiety Disorders are the most common forms of mental illness (1). The National Comorbidity Survey indicated that about 48% of the US population of Americans aged 15 to 54 years had at least one mental illness in their lifetime. Of these, 24.9% reported having an anxiety disorder in their lifetime (1). Anxiety disorders include general anxiety disorder (GAD), panic disorder (PD), and obsessive compulsive disorder (OBC) (2). Of the three disorders, GAD has been shown to be the most prevalent in a population. Patients are required to exhibit at least six symptoms of hyperarousal, vigilance, motor tension, and autonomic hypersensitivity to fit the criteria of GAD in the DSM-IV (2). Pharmacotherapies of anxiety disorders include treatment with benzodiazipines, buspirone, and antidepressants (3). Benzodiazipines have been known for their high efficacy and their fast onset of action. However, they are only recommended for treatment of acute anxiety as long term use of benzodiazipines leads to a wide range of adverse effects including sedation, improper coordination, memory loss, depression, dependence, and potential for abuse (3). The use of antidepressants, venlafaxine, imipramine, and paroxetine, in the treatment of anxiety has been shown to be efficacious; however, these drugs exhibit a delayed onset of action and various adverse effects including nausea, insomnia, jitteriness, restlessness and agitation. As a result, patients on antidepressants usually terminate the treatment before full recovery from anxiety is accomplished (3). The use of TCAs, such as clomipramine, is also efficacious; however, its anticholinergic side effects are so severe that patients also tend to end the treatment before full recovery is reached (1). Studies with buspirone have showed that it also exhibits a more gradual onset of action and may have a lower efficacy than benzodiazipines. Its adverse effects include GI symptoms and dizziness (3).

As a result, a newer anxiolytic agent is required that has similar or better characteristics of efficacy and onset of action as benzodiazipines, exhibits a better profile of safety and no dependence, and indicated for long term treatment of anxiety.

In this chapter, a double blind, placebo controlled, parallel-group study design will be used to test the efficacy and safety of investigational drug XXX versus placebo in patients diagnosed with moderate to severe anxiety. Previous clinical studies have shown that drug XXX is a highly potent serotonin 1A receptor agonist, is not structurally or chemically related to benzodiazipines, has no sedative effect, does not lead to dependence, has a fast onset of action, requires once daily dosing and is eliminated by biotransformation. In this study design, these properties of the drug will be better characterized and its efficacy will be determined. The example of the study design illustrated in this chapter is most applicable to trials for evaluating anxiolytic drugs intended to be used in patients with moderate to severe anxiety as the first line of treatment.

II. PHASE II STUDIES FOR REGISTRATION OF NEW ANXIOLYTIC DRUGS

II.1. Outline of a typical development plan

This study will examine the efficacy and safety of drug XXX in men and women experiencing moderate to severe anxiety. All patients enrolled in the study meeting inclusion/exclusion criteria and that give consent will be randomly assigned to receive one oral daily dose of drug XXX or placebo for 6 months. Efficacy and safety measures are going to be performed at weekly intervals up to week 4, then at 2-week intervals up to week 12, and then at 4-week intervals up to the end of the 6 months. The study will be randomized, double blind, placebo-controlled and patients will come back for two follow-up assessments.

II.2. Long-term studies

II.2.a. Study Objectives

Primary Objectives

a. To compare the efficacy of drug XXX treatment versus placebo in reducing the symptoms of anxiety.

b. To compare the safety of drug XXX treatment versus placebo
Secondary Objectives
a. To determine the onset of the anxiolytic action of drug XXX
b. To determine the duration of the anxiolytic effect of drug XXX
c. To determine the peak anxiolytic effects of drug XXX

II.2.b. Primary Endpoints
Rating scales will be administered to assess the following dependent variables:

a. Structured Clinical Interview for DSM-IV to diagnose patients with Generalized Anxiety Disorder
b. The Hamilton Rating Scale for Anxiety, HAM-A to assess severity of anxiety (score \( \geq 18 \) for patients with severe anxiety and a score of \( \geq 2 \) on the HAM-A item 1 (anxious mood) and item 2 (tension)). It is a simple 14-item five step rating scale. Each item is a group of symptoms that represents one general criterion associated with the disorder. For example, the combination of worries, anticipation of the worst, apprehension, and irritability determine anxious mood. This scale is simple, specific for anxiety measurement, and sensitive to drug effects (4).
c. The Stait-Trait Anxiety, STA-IX to measure subject selection (the Trait) and treatment effects (the Stait). It is a self-evaluation scale that consists of 20 items with a four step severity scale. The Questionnaire asks the patient to indicate how he/she feels at the moment (state) or in general (trait) by selecting “not at all”, “somewhat”, “moderately so”, or “very much so”, with the latter indicating high anxiety (4).

Responders to the anxiolytic drug XXX versus placebo, where a response is defined as:

R = a reduction from baseline (week 1) to week 8 and a similar or a further reduction from baseline to the end of month 6 as measured by the HAM-A, in order to determine the differences in efficacy and safety between the short-term outcome (by week 8) and the long-term outcome (by the end of month 6).

R = a reduction in anxiety measures from baseline to week 8 and a similar or further reduction from baseline to the end of month 6 as measured by the STA-IX scale.

II.2.c. Secondary Endpoints

a. The time of onset of a consistent decrease in anxious mood as measured by the Visual Analogue Anxiety Scale (VAAS) compared to baseline (4).
b. The day during which the greatest reduction in anxious mood is present as measured by the HAM-A.
c. The number of patients that achieve HAM-A \( \leq 10 \) or full recovery.

Adverse effects are measured by changes in blood pressure, heart rate, GI motility, and abnormal laboratory tests (blood and liver).

II.2.d. Study Design
This is a double blind, randomized, placebo-controlled, parallel-group study. The subjects will be administered either a single oral dose of drug XXX or a single oral dose of placebo daily for 6 months. The number of patients receiving drug XXX will equal the number of patients receiving placebo within each group. The anxiolytic effects of drug XXX will be measured using various questionnaires (discussed below). Objective measures, such as blood pressure, will also be obtained. The results will be compared and analyzed between and within groups.

Screening for Eligibility (Visit 1)
Subjects will be interviewed to ensure suitability for study participation a week before the commencement of the study. The following will be required to assess eligibility:
a. Written informed consent.
b. Structured Clinical Interview for DSM-IV to assess psychiatric status and to rule out dependence on psychoactive substances.

c. Current level of anxiety (Hamilton Depression Scale (HAM-A), State-Trait Anxiety (STA-IX)). This is the baseline measure which all upcoming results will be compared against.

d. Brief medical examination (heart rate, blood pressure).

e. Medical history.

f. Review of inclusion/exclusion criteria.

g. Pregnancy test for women.

h. Blood and urine samples to assess liver function, hematology, biochemistry, and to detect the presence of other psychoactive drugs.

i. Participants will be provided with a pager number to be used if they experience any serious side effects and a wallet card containing information about participation in this study.

j. Patients are also given a diary in which they record drug compliance and any side effects that they may experience on a daily basis.

Treatment Phase (Visits 1-10, Months 1-6)

a. Eligible subjects will attend 11 treatment sessions (weekly intervals up to week 4, then at 2-week intervals up to week 12, and then at 4-week intervals up to the end of the 6 months).

b. Medical examination.

c. Medication will be dispensed (enough pills to last for the next visit at once daily dosing).

d. Treatment will consist of 30 to 45 minute sessions with the research assistant.

e. A psychiatrist will be available for consultation, assessment, and treatment as needed (i.e. adverse drug reaction, increases in severity of anxious symptomatology).

f. Review Daily Diary forms on which patients record compliance with medication and any side effects that the patient may be experiencing.

g. At each visit, the HAM-A and the STA-IX, and VAS will be completed and subjects will be interviewed regarding concomitant illness and medication use.

h. Ask patients to return any unused medication in the vial.

i. Blood will be drawn for trough drug concentrations at visits 2, 4, 6, 8, 9, 10, and 11 (weeks 2, 4, 8, 12, 16, 20, 24 after commencing medication).

j. Blood and urine will also be collected at visits 2, 4, 6, 8, 9, 10, and 11 for drug screen, complete biochemistry and hematology analysis.

k. Subjects will be referred to their family physicians either at the end of the 6 months or if a subject decides to terminate participation in the study.

l. Individuals who do not respond to drug XXX will be referred to alternate psychiatric treatment or to their family physicians.

Follow-up visits 11 and 12 (at 8 months and at the 12 months after treatment)

a. Review daily diary.

b. Medical examination.

c. Psychiatrist: examine any increased anxious symptoms, interview patients for concomitant illness, and examine potential adverse reactions.

d. Complete questionnaires: HAM-A, STAI-XI, VAS.

e. Blood and urine collection for drug screen, complete biochemistry, and hematology analyses.

II.2.f. Planned Sample

Refer to Flemming’s Single Stage Procedure Subsection V (planned sample) of the Mood disorders Section II.2.e. in Chapter 20.
II.2.f. Study Population
Male or Female over 18 years of age meeting DSM-IV criteria for GAD and who exhibit moderate to severe anxiety or a Ham-A score of $\geq 18$, and a score of $\geq 2$ on the HAM-A item 1 (anxious mood) and item 2 (tension).

II.2.g. Specific Inclusion Criteria
A subject will be eligible for inclusion in the study only if all of the following criteria apply:

a. Males or females between 19 to 50 years of age.

b. Socially stable.

c. Meet DSM-IV criteria for general anxiety disease.

d. In-patients or out-patients.

e. Non smokers.

II.2.h. Specific Exclusion Criteria
Exclusion criteria must take into account the characteristics of the drug (pharmacokinetics, and pharmacodynamics, drug-drug interactions, and adverse effects). Patients will not be eligible to participate in the study if any of the following criteria apply:

a. If they meet criteria for comorbid anxiety and MDD, or Panic disorder, or Obsessive compulsive disorder.

b. Active suicidal ideation.

c. If they meet criteria for Bipolar Disorder, schizophrenia, schizo-affective or other substance abuse/dependence.

d. Evidence of medical or surgical illness requiring treatment.

e. History of psychoactive drug dependence or a positive urine test for psychoactive drugs

f. Use of medications which may interfere with the study procedures (e.g. SSRIs).

g. Any clinically significant abnormality evident in biochemistry or hematology test results or in urine analysis requiring further investigation.

h. Receiving or will receive other investigational drug during the study.

i. Pregnant or lactating females.

II.2.i. Tools to Assess Endpoints
The tools used to assess efficacy are shown in the following table.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Assessment Tools to Measure Variable</th>
<th>Time of Assessment</th>
</tr>
</thead>
<tbody>
<tr>
<td>Diagnosis of Anxiety</td>
<td>DSM-IV, HAM-A</td>
<td>Visit 1 (baseline)</td>
</tr>
<tr>
<td>Level and Severity of Anxiety</td>
<td>HAM-A, VAAS</td>
<td>Visit 1 (baseline)</td>
</tr>
<tr>
<td>Reduction in symptoms of anxiety</td>
<td>HAM-A, Sta-IX, VAAS</td>
<td>Visits 1-12</td>
</tr>
<tr>
<td>Reduction in Trait Effects</td>
<td>Sta-IX</td>
<td>Visits 1-12</td>
</tr>
<tr>
<td>Reduction in Treatment Effects</td>
<td>Sta-IX</td>
<td>Visits 1-12</td>
</tr>
<tr>
<td>Time of onset of a consistent decrease in anxiety symptoms</td>
<td>HAM-A, VAAS</td>
<td>Visits 1-12</td>
</tr>
<tr>
<td>Time of greatest reduction in symptoms of anxiety</td>
<td>HAM-A</td>
<td>Visits 1-10</td>
</tr>
<tr>
<td>Number of patients achieving HAM-A &lt; 10</td>
<td>HAM-A</td>
<td>Visits 1-10</td>
</tr>
<tr>
<td>Drug Compliance</td>
<td>Patient Daily Diaries and Returned Medication</td>
<td>Visits 1-8</td>
</tr>
</tbody>
</table>
The tools to assess safety are described in Subsection II.2.i. (Tools to Assess Safety) of the Mood Disorders Section of Chapter 20.

II.2.j. Specific criteria for early withdrawal and discontinuation
Refer to Subsection II.2.j. (Specific criteria for early withdrawal and discontinuation) of the Mood Disorders Section of Chapter 20.

II.2.k. Data Analysis Method
Refer to Subsection II.2.k. (Data Analysis Method) of the Mood Disorders Section of Chapter 20.

III. REFERENCES

Chapter 23. Schizophrenic Disorders

Claudio A. Naranjo, M.D. ¹

Lara Chayab ²

¹ Professor
Department of Pharmacology
Psychiatry and Medicine
University of Toronto
Toronto, Ontario
CANADA
and
Head
Neuropsychopharmacology Research Program
Sunnybrook & Women’s College Health Science Centre
Toronto, Ontario
CANADA

² M.Sc. Candidate
Department of Pharmacology
University of Toronto
Toronto, Ontario
CANADA
and
Researcher
Neuropsychopharmacology Research Program
Sunnybrook & Women’s College Health Science Centre
Toronto, Ontario
CANADA
I. INTRODUCTORY REMARKS

Psychotic disorders are personality and thought disorders that are associated with emotional and behavioral impairments (1). Schizophrenia is a psychotic disorder that is accompanied by impairments in speech patterns, inability to process information, delusions, and hallucinations (1). Schizophrenia consists of positive and negative symptoms (1). Increased activity, agitation, delusions, and hallucinations characterize positive symptoms (1). The mechanism by which these symptoms are manifest is through an increase in dopaminergic activity in the brain. Negative symptoms, on the other hand, are characterized by a decrease in activity, loss of pleasure, withdrawal from social interactions, a decrease in dopaminergic activity in the cerebral cortex, and an increase in dopaminergic activity in the striatum (1). The diagnostic criteria of Schizophrenia according to the DSM-IV include that the patient’s illness has to be continuous for at least six months and an occurrence of at least one psychotic phase followed by a residual phase (2). Pharmacological treatments of Schizophrenia include the use of first generation antipsychotics, such as haloperidol, and second generation antipsychotics, such as clozapine (3). First generation antipsychotics are potent antagonists of the dopamine 2 receptors in the striatum. They also block serotonergic, cholinergic, adrenergic, and histaminergic receptors (3). The wide array of affinity of these drugs to various receptors leads to side effects including extrapyramidal effects (sedation), tachycardia, dry mouth, blurred vision, gastrointestinal problems, sexual dysfunction, and orthostatic hypotension, Parkinson’s like syndrome, tardive dyskinesia, dystonia, akathisia, and neuroleptic malignant syndrome (1). These side effects are dose-related unlike the efficacy, which reaches a plateau after a certain dose. All first generation (typical) antipsychotics exhibit the same efficacy for the treatment of positive symptoms of schizophrenia, but their effect on negative symptoms has yet to be determined. Second generation (atypical) antipsychotics were developed for patients who are resistant to treatment with first generation antipsychotics (1). They have been shown in many studies to exhibit higher efficacy and slightly better tolerability (better safety profile) than first generation drugs in the treatment of schizophrenia (1). Many atypical antipsychotics block the dopamine 4 receptors in the cerebral cortex. They also have affinity for the serotonin 2 receptors and the dopamine 2 receptors (1). The better tolerability that atypical drugs exhibit accounts for the lower risk of extrapyramidal symptoms; however, adverse effects on the cholinergic and adrenergic systems still exist (1). As a result, an antipsychotic drug is required that exhibits high efficacy for the treatment of both positive and negative symptoms of schizophrenia. In addition, the antipsychotic drug must exhibit a wide therapeutic index, no extrapyramidal side effects, and very low cholinergic and adrenergic effects.

In this chapter, a double blind, placebo controlled, parallel-group study design will be used to test the efficacy and safety of investigational drug YYY versus placebo in patients diagnosed with schizophrenia. Previous clinical studies have shown that drug YYY is a highly potent dopamine 4 receptor antagonist, has no sedative effect, has a fast onset of action, and requires once daily dosing. In this study design, these properties of the drug will be better characterized and its efficacy and safety will be determined. The example of the study design illustrated in this chapter is most applicable to trials evaluating antipsychotic drugs intended to be used in patients with schizophrenia.

II. PHASE II STUDIES FOR REGISTRATION OF NEW ANTIPSYCHOTIC DRUGS

II.1. Outline of a typical development plan

This study will examine the efficacy and safety of drug YYY in men and women diagnosed with schizophrenia according to the DSM-IV. All patients enrolled in the study meeting inclusion/exclusion criteria and that give consent will be randomly assigned to receive one oral daily dose of drug YYY or placebo for 6 months. A long-term trial has been chosen for this study since the cholinergic and adrenergic side effects usually take 3 to 6 months to appear. Efficacy and safety measures are going to be performed
at weekly intervals up to the end of week 12, and then at 2-week intervals up to week 24. The study will be randomized, double blind, placebo-controlled and patients will come back for two follow-up assessments.

II.2. Long-term studies

II.2.a. Study Objectives

Primary Objectives
a. To compare the efficacy of drug YYY treatment versus placebo in reducing the positive and negative symptoms of schizophrenia.
b. To compare the safety of drug YYY treatment versus placebo

Secondary Objectives
a. To determine the onset of the antipsychotic action of drug YYY
b. To determine the duration of the antipsychotic effect of drug YYY
c. To determine the peak antipsychotic effects of drug YYY

II.2.b. Primary Endpoints

Rating scales will be administered to assess the following dependent variables:

a. Structured Clinical Interview for DSM-IV to diagnose patients with schizophrenia.
b. Brief Psychiatric Rating Scale (BSRS) to determine the severity of the disorder and positive and negative symptoms. It is a 16-item scale, which includes 7 points for severity scale, 5 points for positive symptoms, 2 points for negative symptoms, and 9 general symptom points. Patients with schizophrenia score ≥33 points out of 112 (4). Positive and Negative Syndrome Scale (PANSS) which also determines the severity of schizophrenia and more specifically deals with the positive and negative symptoms associated with schizophrenia. It is a 30-item scale, which includes 7 points that measure positive symptoms, 7 points for negative symptoms, and 16 points for general psychopathology symptom measure. A schizophrenic patient would typically score 91 at beginning of trial (4).
c. Clinical Global Impression Scale (CGI) is used to assess treatment response in psychiatric patients. It is a 3-item scale that measures the severity of the illness (7-point scale), global impairment (7-point scale), and efficacy index (4-point scale). This scale is taken at baseline and repeated after drug exposure in order to compare results and assess efficacy (4).
d. Simpson Angus Scale (SAS) detects any drug induced parkinsonism and extrapyramidal side effects. It evaluates the presence and severity of the symptoms using a 10-item rating scale (4).
e. Barnes Akathisia Scale (BAS) is a 4-item scale that detects the presence and severity of any drug induced akathisia. The scale measures the objective and subjective effects such as restlessness and awareness of restlessness respectively (4).
f. Extrapyramidal Symptom Rating Scale (ESRS) is a 12-item scale that detects the presence of drug induced parkinsonism like symptoms, akathisia, dyskinesia, and dystonia (4).
g. Abnormal Voluntary Movement Scale (AIMS) is a scale that detects the patient’s movements by providing certain positions in which the patients have to rotate his/her body and the psychiatrists assess whether abnormal facial or body movements exist (4).

Responders to the antipsychotic drug YYY versus placebo, where a response is defined as:

a. R = a reduction from baseline (visit 1) during (weeks 1 – 24) and post treatment (2 follow-up sessions) as measured by the BSRS.
b. R = a reduction from baseline (visit 1) during (weeks 1 – 24) and post treatment (2 follow-up sessions) as measured by the PANSS.
c. \( R \) = a reduction in the CGI scale scores on weeks 1-24 compared to baseline will indicate the presence of drug efficacy. The higher a reduction in the score, the more efficacious the drug is considered.

d. \( R \) = an increase in the SAS scale scores on weeks 1-24 compared to baseline will indicate a presence of parkinsonism-like adverse effect.

e. \( R \) = an increase in the BAS scale scores on weeks 1 – 24 compared to baseline will indicate the presence of akathasia.

f. \( R \) = an increase in the ESRS scale scores on weeks 1 – 24 compared to baseline will indicate the presence of extrapyramidal adverse effects.

g. \( R \) = an increase in the AIMS scale scores on weeks 1 – 24 compared to baseline will indicate the presence of abnormal voluntary movements.

II.2.c. Secondary Endpoints

a. The time of onset of a consistent decrease in schizophrenic symptoms as measured by the CGI compared to baseline.

b. The treatment day during which the greatest reduction in the schizophrenic symptoms is present as measured by the PANSS.

c. The number of patients that achieve a BSRS < 33 and a PANSS < 91 by the end of the study period.

d. Time of onset of any adverse effect as measured by the SAS, BAS, ESRS, and AIMS scales.

e. Severity of the adverse effects (if developed) as measured by the SAS, BAS, ESRS, and AIMS scales.

f. Number of subjects that drop out because of the development of adverse effects.

Adverse effects are measured by changes in blood pressure, heart rate, GI motility, and abnormal laboratory tests (blood and liver).

II.2.d. Study Design

This is a double blind, randomized, placebo-controlled, parallel-group study. There will be 2 groups in this study, schizophrenic patients and a healthy control group. Patients in each group will receive either a single dose of drug YYY or a single dose of placebo randomly once daily for 24 weeks. The number of patients receiving drug YYY will equal the number of patients receiving placebo within each group. The antipsychotic effect of drug YYY will be measured using various questionnaires (discussed below). Objective measures, such as blood pressure, will also be obtained. The results will be compared and analyzed between and within groups.

Screening for Eligibility (Visit 1, One week before start of the study)

Subjects will be interviewed to ensure suitability for study participation a week before the commencement of the study. The following will be required to assess eligibility:

a. Written informed consent.

b. Structured Clinical Interview for DSM-IV to assess psychiatric status and to rule out dependence on psychoactive substances.

c. Current level of schizophrenia (Brief Psychiatric Rating Scale (BSRS), Positive and Negative Syndrome Scale (PANSS), Clinical Global Impression Scale (CGI), Simpson Angus Scale (SAS), Barnes Akathisia Scale (BAS), Extrapyramidal Symptom Rating Scale (ESRS), Abnormal Voluntary Movement Scale (AIMS)). This is the baseline measure to which all upcoming results will be compared.

d. Brief medical examination (heart rate, blood pressure).

e. Medical history.

f. Review of inclusion/exclusion criteria.
g. Pregnancy test for women.

h. Blood and urine samples to assess liver function, hematology, biochemistry, and to detect the presence of other psychoactive drugs.

i. Participants will be provided with a pager number to be used if they experience any serious side effects and a wallet card containing information about participation in this study.

j. Patients are also given a diary in which they record drug compliance and any side effects that they may experience on a daily basis.

**Treatment Phase (Visits 2-19, Months 1-6)**

a. Eligible subjects will attend 24 treatment sessions (weekly intervals up to the end of the 6 months).

b. Medical examination.

c. Medication will be dispensed (enough pills to last for the next visit at once daily dosing)

d. Treatment will consist of 60 minutes sessions with the research assistant.

e. A psychiatrist will be available for consultation, assessment, and treatment as needed (i.e. adverse drug reaction, increases in severity of anxious symptomatology).

f. Review Daily Diary forms on which patients record compliance with medication and any side effects that the patient may be experiencing.

g. At each visit, the BSRS, PANSS, SAS, BAS, ESRS, and AIMS will be completed and subjects will be interviewed regarding concomitant illness and medication use.

h. Ask patients to return any unused medication in the vial.

i. Blood will be drawn for trough drug concentrations at 2-week intervals after the start of the study.

j. Blood and urine will also be collected at 2-week intervals for drug screen, complete biochemistry and hematology analysis.

k. Subjects will be referred to their family physicians either at the end of the 6 months or if a subject decides to terminate participation in the study.

l. Individuals who do not respond to drug YYY will be referred to alternate psychiatric treatment or to their family physicians.

**Follow-up visits (Visits 20-21, at 8 months and 12 months after treatment)**

a. Review daily diary.

b. Medical examination.

c. Psychiatrist: examine any increased schizophrenic symptoms, interview patients for concomitant illness, and examine potential adverse reactions.

d. Complete questionnaires: BSRS, PANSS, SAS, BAS, ESRS and AIMS.

e. Blood and urine collection for drug screen, complete biochemistry and hematology analysis.

II.2.e. **Planned Sample**

Refer to Flemming’s Single Stage Procedure Subsection II.2.e. (Planned sample) of the Mood disorders Section in Chapter 20.

II.2.f. **Study Population**

Male or female over 18 years of age meeting DSM-IV criteria for Schizophrenia and score ≥33 on the BSRS and ≥ 91 on the PANSS.

II.2.g. **Specific Inclusion Criteria**

A subject will be eligible for inclusion in the study only if all of the following criteria apply:

a. Males or females between 19 to 50 years of age.

b. Socially stable.
c. Meet DSM-IV criteria for schizophrenia.
d. In-patients or out-patients.
e. Non-smokers.

II.2.h. Specific Exclusion Criteria
Exclusion criteria must take into account the characteristics of the drug (pharmacokinetics, and pharmacodynamics, drug-drug interactions, and adverse effects). Patients will not be eligible to participate in the study if any of the following criteria apply:
   a. If meet criteria for Bipolar Disorder, MDD, Anxiety, or substance abuse/dependence.
   b. Evidence of medical or surgical illness requiring treatment.
   c. History of psychoactive drug dependence or a positive urine test for psychoactive drugs.
   d. Use of medications which may interfere with the study procedures (e.g. SSRIs).
   e. Any clinically significant abnormality evident in biochemistry or hematology test results or in urine analysis requiring further investigation.
   f. Receiving or will receive other investigational drug during the study.
   g. Pregnant or lactating females.

II.2.i. Tools to assess endpoints
The tools used to assess efficacy are shown in the following table.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Assessment Tools to Measure Variable</th>
<th>Time of Assessment</th>
</tr>
</thead>
<tbody>
<tr>
<td>Diagnosis of Schizophrenia</td>
<td>DSM-IV</td>
<td>Visit 1 (baseline)</td>
</tr>
<tr>
<td>Level and Severity of Schizophrenia</td>
<td>BSRS, PANSS, CGI</td>
<td>Visit 1 (baseline)</td>
</tr>
<tr>
<td>Reduction in Schizophrenic Symptoms</td>
<td>BSRS, PANSS, CGI</td>
<td>Visits 2-21</td>
</tr>
<tr>
<td>Reduction in Positive Symptoms</td>
<td>BSRS, PANSS</td>
<td>Visits 2-21</td>
</tr>
<tr>
<td>Reduction in Negative Symptoms</td>
<td>BSRS, PANSS</td>
<td>Visits 2-21</td>
</tr>
<tr>
<td>Induction of Extrapyramidal Adverse Effects</td>
<td>SAS, ESRS</td>
<td>Visits 2-21</td>
</tr>
<tr>
<td>Time of induction and Severity Extrapyramidal AE</td>
<td>SAS, ESRS</td>
<td>Visits 2-21</td>
</tr>
<tr>
<td>Induction of Parkinsonism-like Symptoms</td>
<td>SAS</td>
<td>Visits 2-21</td>
</tr>
<tr>
<td>Time of Induction and Severity of Parkinsonism-like Symptoms</td>
<td>SAS</td>
<td>Visits 2-21</td>
</tr>
<tr>
<td>Induction of Akathisia</td>
<td>BAS</td>
<td>Visits 2-21</td>
</tr>
<tr>
<td>Time of Induction and Severity of Akathisia</td>
<td>BAS</td>
<td>Visits 2-21</td>
</tr>
<tr>
<td>Induction of Involuntary Movements</td>
<td>AIMS</td>
<td>Visits 2-21</td>
</tr>
<tr>
<td>Time of Induction and Severity of Involuntary Movements</td>
<td>AIMS</td>
<td>Visits 2-21</td>
</tr>
<tr>
<td>Time of onset of a consistent decrease in schizophrenic symptoms</td>
<td>BSRS, PANSS, CGI</td>
<td>Visits 2-21</td>
</tr>
<tr>
<td>Time of greatest reduction of Schizophrenic symptoms</td>
<td>BSRS, PANSS, CGI</td>
<td>Visits 2-21</td>
</tr>
<tr>
<td>Number of patients achieving BSRS &gt; 33</td>
<td>BSRS</td>
<td>Visits 2-21</td>
</tr>
<tr>
<td>Number of patients achieving PANSS &gt; 91</td>
<td>PANSS</td>
<td>Visits 2-21</td>
</tr>
<tr>
<td>Drug Compliance</td>
<td>Patient Daily Diaries andReturned Medication</td>
<td>Visits 2-19</td>
</tr>
</tbody>
</table>
The tools to assess safety are described in Subsection II.2.i. (Tools to Assess Safety) of the Mood Disorders Section of Chapter 20.

II.2.j. Specific criteria for early withdrawal and discontinuation
Refer to Subsection II.2.j. (Specific criteria for early withdrawal and discontinuation) of the Mood Disorders Section of Chapter 20.

II.2.k. Data Analysis Method
Refer to Subsection II.2.k. (Data Analysis Method) of the Mood Disorders Section of Chapter 20.

III. REFERENCES

Chapter 24. Alcohol and Nicotine Addiction

Claudio A. Naranjo, M.D. 1

Lara Chayab 2

1 Professor
Department of Pharmacology
Psychiatry and Medicine
University of Toronto
Toronto, Ontario
CANADA
and
Head
Neuropsychopharmacology Research Program
Sunnybrook & Women’s College Health Science Centre
Toronto, Ontario
CANADA

2 M.Sc. Candidate
Department of Pharmacology
University of Toronto
Toronto, Ontario
CANADA
and
Researcher
Neuropsychopharmacology Research Program
Sunnybrook & Women’s College Health Science Centre
Toronto, Ontario
CANADA
I. INTRODUCTORY REMARKS

Alcohol, opioids, nicotine, and psychostimulants have different chemical structures; however, they seem to exert their actions via similar neurochemical pathways in the brain, which lead to addiction. Current pharmacotherapy approaches available for the treatment of alcohol and drug addiction aim at minimizing symptoms of acute abstinence and the risk of relapse. Alcoholism is a complex disorder exhibiting multiple symptoms. It is often co-morbid with Major Depressive Disorders, antisocial personality, or anxiety. According to the American Psychiatric Association (DSM-IV), individuals must meet three of the following criteria during a 12-month period for a diagnosis of alcohol dependence: a) Tolerance to alcohol, increase amounts of alcohol consumption to achieve same effects, b) Signs or symptoms of alcohol withdrawal, c) Attempts to cut down are unsuccessful, d) Long periods of time spent in obtaining alcohol, alcohol consumption, and hangovers, e) Impaired social and work activities due to alcohol consumption, and f) Alcohol consumption is not decreased even if it leads to adverse effects physically and psychologically (1).

Pharmacological treatment of alcohol dependence include agents that minimize the positive reinforcing effects of alcohol (such as naltrexone) and other agents used to relieve withdrawal symptoms and promote abstinence (such as Sedatives and Disulfiram) (2). Naltrexone is a mu-opioid and a delta-opioid receptor antagonist. It functions by blocking the binding of the endogenous opioid, beta-endorphin, to the mu-opioid receptor, which leads to the alleviation of positive effects (euphoria) induced by alcohol intake (3). It is administered orally, and taken 3 times/week at 100-150mg. It is generally safe, with no known interactions caused by alcohol intake and no withdrawal symptoms after drug discontinuation (3). Studies have shown that naltrexone does not lead to complete abstinence from drinking, however, it may cause a reduction in the amount of alcohol intake and a better control over drinking behaviors (4). Acamprosate exhibits a structure similar to the neurotransmitters GABA and glutamate (5). It is thought to work by stabilizing the neurotransmitter balance seen in alcohol dependent people (5). Several trials have shown that acamprosate is efficacious in the treatment of alcohol dependence and is well tolerated (5). However, approval of acamprasate is still pending in many countries. As a result, there is a need for pharmacological agents that treat alcohol dependence, maintain abstinence from alcohol, and do not exhibit adverse effects.

Persistence of cigarette smoking leads to nicotine addiction. Smoking is the leading cause of death in North America, implicated in one of every five deaths (4). Unaided attempts of smoking cessation are successful in only 5% of people who attempt to quit (4). Most pharmacological agents that are available for smoking cessation are nicotine replacement agents such as nicorette (a chewing gum formulation that contains 2 mg of nicotine), and nicotine patch (6). In addition, some investigational drugs for smoking cessation include nicotine inhalers, mecamylamine, a nicotine receptor antagonist, antidepressants, clonidine, and airway sensory replacement (6). Nicorette produces adverse effects that include bad taste, difficulty with chewing, and stomach upset (6). The nicotine patch is better tolerated; however, it may lead to skin irritation and allergies in patients (6). As a result, a pharmacological agent is required that functions to cease cigarette smoking in individuals that are mild to heavy smokers, decreases craving, has no side effects and prevents relapse.

In this chapter, two separate double blind, placebo-controlled, parallel-group study designs will be carried out to test the efficacy and safety of investigational drug AAA (in study-1 for patients with moderate to severe depression, and investigational drug SSS (in study-2 for patients with nicotine addiction. Previous clinical studies have shown that drug AAA is a highly purely selective mu-opioid receptor antagonist, has a half-life of 24 hrs, and is administered 3 times/week. In addition, drug AAA has no potential for alcohol interactions, has mild adverse effects, and has no potential for withdrawal symptoms after drug discontinuation. With the design of the present study, the properties of drug AAA will be better characterized and its efficacy will be determined. Drug SSS is a highly selective nicotine receptor
antagonist that blocks the physiological, behavioral, and reinforcing effects of nicotine. It has a half-life of 12 hours, which accounts for its twice daily dosing, and it is well tolerated due to its mild adverse effects. The properties of Drug SSS will be further investigated and its efficacy will be determined with the present experimental design. The experimental design of both study is most applicable to trials intended to be used in patients as a first line of treatment.

II. PHASE II STUDIES FOR REGISTRATION OF NEW DRUGS FOR THE TREATMENT OF ALCOHOL AND NICOTINE DEPENDENCE

II.1 Outline of a typical development plan

The two studies will examine the efficacy and safety of drug AAA in men and women experiencing moderate to severe alcohol dependence and drug SSS in men and women experiencing moderate to severe nicotine addiction. All patients enrolled in study 1 meeting inclusion/exclusion criteria for alcohol dependence and that give consent will be randomly assigned to receive one oral dose of drug AAA 3 times/week or placebo for 3 months. In addition, patients enrolled in study 2 meeting inclusion/exclusion criteria for nicotine addiction will be randomly assigned two oral daily doses (every 12 hours) of drug SSS or placebo for 3 months. Efficacy and safety measures are going to be performed at weekly intervals up to the end of the third month. The study will be a randomized, double blind, placebo-controlled and patients will come back for two follow-up assessments.

II.2. Short-term studies

II.2.a. Study Objectives

Primary objectives of study 1
a. To compare the efficacy of drug AAA treatment versus placebo in the abstinence from alcohol in alcohol dependent patients.
b. To compare the safety of drug AAA treatment versus placebo.

Secondary objectives for study 1
a. To determine the onset of a decrease in alcohol intake by drug AAA.
b. To determine the duration of the decrease in alcohol intake by drug AAA.
c. To determine the time at which a complete abstinence from alcohol takes place.
d. To determine the time at which a consistent abstinence from alcohol takes place.
e. To determine if patients will relapse after abstinence.
f. The time at which patients start drinking again if relapse took place.

Primary objectives of study 2
a. To compare the efficacy of drug SSS treatment versus placebo in the abstinence from cigarette smoking in patients with nicotine addiction.
b. To compare the safety of drug SSS treatment versus placebo.

Secondary objectives for study 2
a. To determine the onset of a decrease in cigarette smoking by drug AAA.
b. To determine the duration of the decrease in cigarette smoking by drug AAA.
c. To determine the time at which a complete abstinence from cigarette smoking takes place.
d. To determine the time at which a consistent abstinence from cigarette smoking takes place.
e. To determine if patients will relapse after abstinence.
f. The time at which patients start drinking again if relapse took place.
II.2.b. Primary Endpoints

Primary endpoints of study 1
Rating scales will be administered to assess the following dependent variables:

a. Structured Clinical Interview for DSM-IV to diagnose patients with alcohol dependence
b. Alcohol Dependence Scale (ADS): The ADS provides a quantitative measure of the severity of alcohol dependence consistent with the concept of the alcohol dependence syndrome. Its 25 items cover alcohol withdrawal symptoms, impaired control over drinking, awareness of a compulsion to drink, increased tolerance to alcohol, and salience of drink-seeking behaviour. Alcohol dependent patients entering the study must score $\geq 22$ on the ADS, while healthy individuals must score $\leq 2$ (7).

c. Michigan Alcohol Screening Test (MAST): Consisting of 25 questions, the MAST serves to uncover the problems the individual is experiencing as a result of his/her alcohol dependence. Because of the seemingly neutrality of some of the questions, it is easier to extract pertinent information about one's affliction which that person might have been otherwise reluctant to admit. Alcoholic dependent patients entering the study must score $\geq 6$ while healthy patients in the control group must score $\leq 2$ (7).

d. Clinical Institute Withdrawal Assessment for Alcohol (revised) (CIWA): This 10-item scale is used to measure the severity of alcohol withdrawal symptoms. It is important for our results that subjects are not undergoing withdrawal while being tested. All patients entering the study must score $\leq 15$ for no signs of withdrawal detected.

Responders to drug AAA versus placebo, where a response is defined as:

a. $R =$ a reduction from baseline (visit 1) on weeks 1-12 during and post treatment as measured by the ADS.
b. $R =$ a score of $\leq 2$ as measured by the MAST during (weeks 1-12) and post treatment (follow-up sessions) with drug AAA.
c. $R \leq 15$ on the CIWA for no signs of withdrawal detected.

Primary endpoints of study 2
Rating scales will be administered to assess the following dependent variables:

a. Structured Clinical Interview for DSM-IV to diagnose patients with nicotine dependence.

b. Fagerstrom Test for Nicotine Dependence (FTND) is used to assess tobacco dependence. The questionnaire contains items that determine the number of cigarettes smoked per day, the time to the first cigarette after awakening, and the difficulty of restraining from smoking when strongly advised to (ill). Patients with moderate to severe nicotine dependence typically score $\geq 6$. Non-smokers (in the control group) should have 0 points (7).

Responders to drug SSS versus placebo, where a response is defined as:

a. $R =$ any reduction from baseline (visit 1) on weeks 1-12 during and post treatment as measured by the FTND.

II.2.c. Secondary endpoints

Secondary endpoints for study 1

a. The first day during which a reduction in alcohol intake is seen as measured by ADS and as reported by the patient daily diaries.

b. The treatment day during which the greatest reduction in alcohol intake is present as measured by the ADS.

c. The day at which a complete abstinence from alcohol takes place as measured by the ADS and as reported by the daily diaries.

d. The amount of days during which a consistent abstinence from alcohol takes place.
e. The number of patients that achieve $\text{ADS} \leq 2$.
f. The day at which patients start drinking alcohol again if relapse occurs.

Secondary endpoints for study 2
a. The first day during which a reduction in cigarette smoking is seen as measured by FTND and as reported by the patient daily diaries.
b. The treatment day during which the greatest reduction in cigarette smoking is present as measured by the FTND.
c. The day at which a complete abstinence from cigarette smoking takes place as measured by the FTND and as reported by the daily diaries.
d. The amount of days during which a consistent abstinence from cigarette smoking takes place.
e. The number of patients that achieve $\text{FTND} = 0$.
f. The day at which patients start smoking again if relapse occurs.

Adverse effects are measured by changes in blood pressure, heart rate, GI motility, and abnormal laboratory tests (blood and liver).

II.2.d. Study Design
Both studies are double blind, randomized, placebo-controlled, parallel-group study. There will be 2 groups in study 1, a moderate to severely alcohol dependent group, and a healthy control group. Patients in each group will receive either a single dose of drug AAA or a single dose of placebo randomly once daily, 3 times/week, for 12 weeks. In addition, there will also be 2 groups in study 2, a moderate to severely nicotine dependent group, and a healthy (non-smokers) control group. Patients in each group will receive either drug SSS or placebo twice a day for 12 weeks. The number of patients receiving drug AAA/SSS will equal the number of patients receiving placebo within each group. The effects of drug AAA/SSS will be measured using various questionnaires (discussed below). Objective measures, such as blood pressure, will also be obtained. The results will be compared and analyzed between and within groups.

Screening for Eligibility (Visit 1)
Subjects will be interviewed to ensure suitability for study participation a week before the commencement of the study. The following will be required to assess eligibility:
a. Written informed consent.
b. Structured Clinical Interview for DSM-IV to assess dependence on psychoactive substances (alcohol or nicotine).
c. Current level of alcohol dependence (moderate to severe alcohol dependent patients must have a score of $\geq 22$ on the ADS while healthy patients in the control group must score ADS $\leq 2$). Current level of nicotine dependence (moderate to severe nicotine dependent patients must score $\geq 3$ on the FTND while non-smokers in the control group must have no FTND score. This is the baseline measure to which all upcoming results will be compared against.
d. Brief medical examination (heart rate, blood pressure).
e. Medical history.
f. Review of inclusion/exclusion criteria.
g. Pregnancy test for women.
h. Blood and urine samples to assess liver function, hematology, biochemistry, and to detect the presence of other psychoactive drugs.
i. Participants will be provided with a pager number to be used if they experience any serious side effects and a wallet card containing information about participation in this study.
j. Patients are also given a diary in which they record drug compliance and any side effects that they may experience on a daily basis.
Treatment Phase (Visits 2 -7, Weeks 1-12)

a. Eligible subjects will attend six treatment sessions (one every two weeks).
b. Medical examination.
c. Medication will be dispensed (enough pills for two weeks).
d. Treatment will take place at 2-week intervals consisting of 30 to 45 minute sessions with the research assistant.
e. A psychiatrist will be available for consultation, assessment, and treatment, as needed (i.e. adverse drug reaction, any withdrawal symptoms).
f. Review Daily Diary forms on which patients record compliance with medication.
g. Study 1: At each visit, the ADS, MAST, and CIWA will be completed and subjects will be interviewed regarding concomitant illness and medication use.
h. Study 2: At each visit, the FTND will be completed and subjects will be interviewed regarding concomitant illness and medication use.
i. Ask patients to return any unused medication in the vial.
j. Blood will be drawn for trough drug concentrations at visits 3, 5 and 7 (4, 8 and 12 weeks after commencing medication).
k. Blood and urine will also be collected at visits 4, 6 and 8 for drug screen, complete biochemistry and hematology analyses.
l. Subjects will be referred to their family physicians either at the end of the 12 week study or if a subject decides to terminate participation in the study.
m. Individuals who do not respond to drug AAA/SSS will be referred to alternate psychiatric treatment or to their family physicians.

Follow-up visits (Visits 8-9, at 3 months and 6 months after treatment)

a. Review daily diary.
b. Medical examination.
c. Psychiatrist: examine any increased alcohol/nicotine dependence, interview patients for concomitant illness, and examine potential adverse reactions.
d. Study 1: Complete questionnaires: ADS, MAST, and CIWA to check for increase/decrease/relapse to alcohol dependence.
e. Study 2: Complete questionnaires: FTND to check for increase/decrease/relapse to nicotine dependence.
f. Blood and urine collection for drug screen, complete biochemistry, and hematology analyses.

II.2.e. Planned Sample
Refer to Flemming’s Single Stage Procedure Subsection II.2.e. (planned sample) of the Mood disorders Section in Chapter 20.

II.2.f. Study Population
Males or Females over 18 years of age meeting DSM-IV criteria for alcohol/nicotine dependence and exhibit moderate to severe alcohol dependence or an ADS score of ≥22/ moderate to severe nicotine dependence or an FTND ≥3.

II.2.g. Specific Inclusion Criteria
A subject will be eligible for inclusion in both studies only if all of the following criteria apply:

a. Males or females between 19 to 50 years of age.
b. Socially stable.
c. Meet DSM-IV criteria for alcohol/nicotine dependence.
d. In-patients or out-patients.
e. Non-smokers for study 1.
II.2.h. Specific Exclusion Criteria
Exclusion criteria must take into account the characteristics of the drug (pharmacokinetics, and pharmacodynamics, drug-drug interactions, and adverse effects). Patients will not be eligible to participate in the study if one of the following criteria apply:

a. Meet criteria for MDD, Anxiety, Bipolar disorder, Schizophrenia, Schizo-affective or other substance abuse/dependence (other than alcohol and nicotine).
b. Evidence of medical or surgical illness requiring treatment.
c. History of psychoactive drug dependence (other than alcohol and nicotine) or a positive urine test for psychoactive drugs (other than alcohol and nicotine).
d. Use of medications which may interfere with the study procedures (e.g. SSRIs).
e. Any clinically significant abnormality evident in biochemistry or hematology test results or in urine analysis requiring further investigation.
f. Receiving or will receive other investigational drug during the study.
g. Pregnant or lactating females.

II.2.i. Tools to assess
Tools to assess efficacy in alcohol and nicotine dependence are shown in tables 1 and 2.

<table>
<thead>
<tr>
<th>Table 1. Alcohol Dependence</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Variable</strong></td>
</tr>
<tr>
<td>Diagnosis of Alcohol Dependence (AD)</td>
</tr>
<tr>
<td>Level and Severity of AD</td>
</tr>
<tr>
<td>Reduction in AD</td>
</tr>
<tr>
<td>Reduction in impaired control over drinking</td>
</tr>
<tr>
<td>Reduction in drinking-seeking behavior</td>
</tr>
<tr>
<td>Increase in tolerance to drinking</td>
</tr>
<tr>
<td>Severity of Alcohol Withdrawal</td>
</tr>
<tr>
<td>Time of onset of a consistent decrease in AD</td>
</tr>
<tr>
<td>Time of greatest reduction of AD</td>
</tr>
<tr>
<td>Number of patients achieving ADS &lt; 2</td>
</tr>
<tr>
<td>Number of patients achieving MAST &lt; 2</td>
</tr>
<tr>
<td>Drug Compliance</td>
</tr>
</tbody>
</table>

The tools to assess safety are described in Subsection II.2.i. (Tools to Assess Safety) of the Mood Disorders Section of Chapter 20.

II.2.j. Specific criteria for early withdrawal and discontinuation
Refer to Subsection II.2.j. (Specific criteria for early withdrawal and discontinuation) of the Mood Disorders Section of Chapter 20.

II.2.k. Data Analysis Method
Refer to Subsection II.2.k. (Data Analysis Method) of the Mood Disorders Section of Chapter 20.
### Table 2. Nicotine Dependence

<table>
<thead>
<tr>
<th>Variable</th>
<th>Assessment Tools to Measure Variable</th>
<th>Time of Assessment</th>
</tr>
</thead>
<tbody>
<tr>
<td>Diagnosis of Nicotine Dependence (ND)</td>
<td>DSM-IV</td>
<td>Visit 1 (baseline)</td>
</tr>
<tr>
<td>Level and Severity of ND</td>
<td>FTND</td>
<td>Visit 1 (baseline)</td>
</tr>
<tr>
<td>Reduction in ND</td>
<td>FTND</td>
<td>Visits 2-9</td>
</tr>
<tr>
<td>Reduction in the number of cigarettes smoked per day</td>
<td>FTND</td>
<td>Visits 2-9</td>
</tr>
<tr>
<td>Reduction in the time to first cigarette after awakening</td>
<td>FTND</td>
<td>Visits 2-9</td>
</tr>
<tr>
<td>Time of onset of a consistent decrease in ND</td>
<td>FTND</td>
<td>Visits 2-9</td>
</tr>
<tr>
<td>Time of greatest reduction in ND</td>
<td>FTND</td>
<td>Visits 2-9</td>
</tr>
<tr>
<td>Number of Patients achieving FTND = 0</td>
<td>FTND</td>
<td>Visits 2-9</td>
</tr>
<tr>
<td>Drug Compliance</td>
<td>Patient’s Daily Diaries and Returned Medication</td>
<td>Visits 2-7</td>
</tr>
</tbody>
</table>

### III. REFERENCES

Pharmacological Research in Joint Disorders

Chapter 25. Osteoarthritis/arthrosis Short Term Studies

Chapter 26. Osteoarthritis/arthrosis Long Term Studies

Chapter 27. Rheumatoid Arthritis
Chapter 25. Osteoarthritis/arthrosis Short Term Studies: Pain and Function Improvement

Jean-Pierre Raynauld, M.D. 1

Johanne Martel-Pelletier, Ph.D. 1

Boulos Haraoui, M.D. 1

Jean-Pierre Pelletier, M.D. 1

1 Osteoarthritis Research Unit
Centre hospitalier de l'Université de Montréal (CHUM)
Hôpital Notre-Dame
Département de Medicine
Université de Montréal
Montréal, Québec
CANADA
I. INTRODUCTORY REMARKS

Osteoarthritis (OA) is the most common form of arthritis and is a leading cause of disability. More than 75% of those over age seventy exhibit radiographically detectable changes consistent with osteoarthritis. About 40-60% of subjects with radiological OA changes suffer from clinical symptoms such as pain, joint stiffness, and joint deformities.

Patients with OA have pain that typically worsens with weight bearing and activity and improves with rest, as well as morning stiffness, gelling of the involved joint after periods of inactivity, and limited joint motion. As OA progresses, pain at rest can also be present. With a few exceptions, the causes of OA are not known so that the main goals of therapy are pain relief and improved physical and social function.

Pharmacologic therapy of OA typically begins with analgesics such as acetaminophen in doses up to 4 g/day, progresses to low dose nonsteroidal anti-inflammatory drugs (NSAIDs), and then to full dose NSAIDs (including COX-2 selective inhibitors). NSAIDs, while useful, have a ceiling affect and can be limited in their use because of their side effects, particularly those affecting the gastrointestinal tract, liver, and kidney; the risks of which increase with advanced age.

The primary parameter of this study is the proportion of subjects who achieve adequate pain control (% with moderate, good, or excellent pain control) during 56 days of a new DRUG X. This parameter will be assessed with a pain control assessment performed at study entry, at each telephone contact, and at each visit. Among secondary parameters are the effect on pain intensity, quality of life, functionality, and global assessment of change.

II. RATIONALE FOR STUDY DESIGN

This trial is an observational, therapeutic use study investigating the effect of DRUG X treatment on pain control in subjects with moderate to severe pain due to OA of the hip or knee that is inadequately controlled with acetaminophen or a traditional NSAID.

DRUG X has been demonstrated to be safe and efficacious in chronic non-cancer pain in several randomised clinical trials. Moreover, DRUG X has been demonstrated to be preferred over several NSAIDs with the main reason being that better pain control was achieved. DRUG X was also associated with a better safety profile. The effect of 8 weeks treatment with DRUG X on pain control, quality of life, and functionality has not been previously investigated in a clinical study of subjects with moderate to severe OA pain of the hip or knee that is inadequately controlled with NSAIDs.

II.1. Outline of a Typical Development Plan

Multi-center, randomized, double-blind, placebo-controlled, parallel group study with mild to moderate primary knee (or hip) OA fulfilling the American College of Rheumatology (ACR) criteria (see Appendix A), who have been completely withdrawn from their previous analgesics or anti-inflammatory medications or have been newly diagnosed with mild to moderate primary knee (or hip) OA and who are not currently taking any analgesics or anti-inflammatory medications.

II.2. Short-Term Studies: Pain and Function Improvement

II.2.A. Objectives
The objective of this study is to investigate the short-term effect of a new DRUG X at a dose of Y mg compared to placebo, on pain control in subjects with moderate to severe pain due to OA according to the ACR (see Appendix A) of hip or knee that is inadequately controlled with simple analgesics or NSAIDs.
II.2.B. Primary endpoints
To determine the proportion of subjects with symptomatic osteoarthritis of the hip or knee who achieve adequate pain control (% of subjects with moderate, good, or excellent pain control) during 56 days of treatment with DRUG X.

II.2.C. Secondary endpoints
To compare the scores from the Numerical Pain Intensity Rating scale, WOMAC Osteoarthritis Index questionnaire and Acute SF-36 Health Survey after 56 days of treatment with DRUG X to baseline. Physician and Subject Global Impression of Change scales and Subject Treatment Assessment questionnaire will be done at the Final Visit.

II.2.D. Study design
This is a placebo-controlled, multicentre study.

Eligible subjects will undergo screening procedures. Subjects must show evidence of symptomatic hip or knee OA (ACR Functional Class ≥ grade 2) and meet the ACR hip or knee OA criteria, have “poor” or “very poor” pain control (on a five-point scale of excellent, good, moderate, poor, or very poor), and have at least moderate to severe pain demonstrated as a pain score ≥ 5 on a numerical pain intensity rating scale of 0 to 10 (with “0” representing no pain and “10” representing worst possible pain).

The duration of study treatment is 56 days. At Visit 1 (Study Entry) subjects will start on DRUG X at a dosage Y and will remain on this dose for the trial period.

During each study visit subjects will be required to complete a Pain Control Assessment indicating the amount of pain control experienced that day, a Western Ontario and McMaster Universities (WOMAC) Osteoarthritis Index and an SF-36 Acute Health Survey (Appendix B). Physician and Subject Global Impression of Change scales (7-point scale, see Efficacy Evaluations section) will be conducted at Visit 4. In the event that a subject withdraws early, all Visit 4 procedures will be performed.

At Visit 1 (Study Entry) subjects will start on DRUG X. Supplementary analgesic medication consisting of acetaminophen, 500 mg tablets, will be allowed during the study on an as needed basis, provided the total daily dose of acetaminophen does not exceed 4 g (8 tablets).

Eligible subjects who are enrolled into the study will be randomly allocated to be treated with DRUG X or a placebo for 56 days. Insufficient analgesia will be determined by the investigator, using his or her clinical judgement and taking into account the subject’s level of pain severity, level of pain control, use of supplementary acetaminophen, 500 mg tablets, and individual response and tolerance to the dose.

Subjects will be provided with acetaminophen, 500 mg tablets, as supplementary analgesic medication for any additional pain in the target OA hip or knee joint and will be taken as needed throughout the study (provided the total daily dose of acetaminophen does not exceed 4 g or 8 tablets). Subject use of acetaminophen, 500 mg tablets will be recorded in the Patient diary on a daily basis (Appendix C).

Concomitant analgesic opioid medication is NOT allowed during the course of the study. Weak opioid medication must be discontinued at the time of study entry.

Inhaled steroids for asthma or topical corticosteroid preparations for minor dermatological use will be allowed during the study.

Alcoholic beverages and sedating antihistamines may also produce additive depressant effects and should be used with caution.
ASA (acetylsalicylic acid) for cardiac prophylaxis, up to 325 mg/day, will be allowed during the study and should be used with caution with concomitant use of acetaminophen.

All medications (prescriptions or over-the-counter (OTC) medications, including supplements or nutraceuticals) and medical procedures ongoing in the week preceding study entry that are continued at the start of the study or are started during the trial and are different from the trial medication, must be documented on the Concomitant Therapy Form of the CRF. If any medication or medical procedure is started, stopped, or if the dose or frequency is modified, this must also be documented on the CRF. The sponsor must be notified in advance (or as soon as possible thereafter) of any instances in which prohibited therapies are administered.

For any concomitant therapy given as a treatment for a new condition or a worsening of an existing condition, the condition must be documented on the Adverse Event Form of the CRF.

### III.2.F. Planned sample

Approximately 80 subjects will be required to be able to estimate the proportion of subjects achieving "excellent", "good" or "moderate" pain control within \( \pm 11\% \) with DRUG X compared to placebo with 95% confidence. Because only Visit 2 efficacy response is required for the evaluable population, no contingency is built in the required sample size.

Approximately 80 subjects who have a history of symptomatic OA of the hip or knee with chronic pain for at least 3 months, who have been on a stable daily dose of acetaminophen for at least two weeks prior to Study Entry, and who have uncontrolled pain, will be enrolled into the study.

Approximately 80 subjects from 10 sites will be screened and enrolled in the study. Each site will enrol approximately X subjects. Subjects are OA subjects, suffering with chronic hip or knee pain for at least 3 months (for at least 20 days of each month), and who are not hospitalised. The target joint selected will be an OA hip or knee joint that causes the most pain to the subject. In case of pain of equal severity in hip and knee, one target joint must be selected.

### III.2.G. Study population

Potential study subjects must have a history of symptomatic osteoarthritis of the hip or knee with chronic pain for at least 3 months and must have been on a stable daily dose of acetaminophen at least 2 weeks prior to the study. The study will be explained to subjects and informed consent will be obtained.

### III.2.H. Specific Inclusion Criteria

Subjects must satisfy the ALL of the following inclusion criteria before entering the study:

a. Male or female of ages \( \geq 40 \).

b. Must be in generally good health as confirmed by medical and previous medication history, and baseline physical examination including vital signs.

c. Female subjects must be postmenopausal for at least 2 years, surgically sterile, or practising an effective method of birth control prior to entry and throughout the study, and have a negative urine pregnancy test at the baseline visit. The subject may continue in the study using abstinence as a form of birth control provided that she is completely abstinent, has a negative urine pregnancy test prior to study entry and at the final visit or upon termination (if the subject discontinues the trial early). It must be documented in the medical notes that the subject has been counselled about the birth control and the risks of becoming pregnant.

d. Symptomatic OA of the target hip or knee joint as evidenced by hip or knee pain for at least 3 months (for at least 20 days of each month) and osteophytes confirmed by an x-ray taken within the last two years and who must meet the OA hip or knee criteria of the American College of Rheumatology (Appendix A).
e. After a full explanation of the study, subjects must understand the nature of the study and sign the informed consent form to participate.

f. Subjects with moderate to severe pain of the target OA hip or knee joint whose pain is not adequately controlled with an NSAID. This will be defined as subjects with a pain control assessment of “poor” or “very poor” (on a five-point scale: excellent, good, moderate, poor, or very poor) and a mean pain score $\geq 5$ (on a numerical pain intensity rating scale of 0-10) at Baseline.

III.2.1. Specific Exclusion Criteria
Potential subjects who meet any ONE of the following exclusion criteria will NOT be eligible to participate in the study:

a. Subjects who have previously failed on DRUG X therapy or those who previously have discontinued DRUG X due to adverse events.

b. Subjects who have received treatment with a strong opioid (e.g. morphine, hydromorphone, methadone, long-acting oxycodone, oxymorphone, levorphanol, heroin, etc.) in the 4 weeks preceding study entry. Subjects cannot take strong opioids during the study.

c. Subjects for whom a treatment is planned within the study period that could alter the degree or nature of pain (e.g. arthroscopic techniques, osteotomy, joint replacement surgery, etc.).

d. Subjects who are experiencing another type of continuous pain that is more severe in intensity in comparison with the OA target joint pain (e.g. low back pain, fibromyalgia, ankylosing spondylitis, etc.).

e. Subjects who have had target joint intra- or periarticular corticosteroid injections within 6 weeks of study entry or hyaluronic injections within 6 months of study entry. Injections are not allowed during the study. Subjects cannot have had arthrosynthesis within 4 weeks or arthroscopic techniques (e.g. joint débridement, abrasion, arthroplasty, chondral holes, etc.) within 3 months prior to the study or during the study.

f. Subjects taking glucosamine will not be eligible unless they have been on a stable dose for greater than 2 months preceding study entry. If subjects were taking a stable dose for at least 2 months prior to the study, the dosage should remain constant throughout the study. Glucosamine cannot be started at anytime during the study.

g. Subjects taking NSAIDs, COX-2 selective inhibitors, or steroidal drugs for at least 4 weeks before study entry may continue these medications during the study; however, they must have been taking a stable dose (consistent daily milligram dose $\pm 25\%$) for at least 2 weeks before study entry and the dosage must be kept constant throughout the study. If these medications were started within the 4 weeks preceding the study, the subject will be excluded, but can be rescreened at a later time. These medications cannot be started at anytime during the study.

h. Subjects who have had major surgery in the 3 months preceding the study.

i. Subjects with a significant psychiatric disorder (including major depression) or subjects receiving anti-psychotic medication.

j. Subjects who have taken sedatives, hypnotics, phenothiazines, anticonvulsants, tranquilizers or muscle relaxants two weeks preceding study entry. These medications cannot be started during the study.

k. Subjects who are taking tricyclic antidepressants if not expected to remain on a stable dose of these medications for the duration of the study. These medications cannot be started during the study.

l. Subjects who have applied topical analgesic preparations to the target joint and/or taken general anaesthetics in the one week preceding study entry. These medications cannot be started during the study.

m. Subjects with documented or suspected history of alcohol or drug abuse, or who have a documented or suspected history of an addictive personality.

n. Subjects who have started any form of physiotherapy, acupuncture, TENS, massage or active physical therapy within the 4 weeks preceding study entry. Such therapies can continue if they were started more than 4 weeks before the start of the study and if they continue at the same frequency of administration throughout the study. Any such therapies cannot be started during the study.
o. Female subjects who are breast-feeding.
p. Subjects known to have any of the following:
   • significantly abnormal renal or hepatic function;
   • any disease or condition that compromises the function of those body systems that could
     result in altered absorption, excess accumulation, or impaired metabolism or excretion of the
     test medications;
   • a life-threatening disease (e.g. AIDS, malignant disease, etc.) that would preclude completion
     of study or interfere with protocol compliance;
   • any condition that in the investigator’s judgement precludes participation in the study.
q. Subjects who have received an investigational drug or have used an investigational device in the
   30 days preceding study entry.

II.2.J. Tools for assessing endpoints
Clinic assessments will be completed at four different time points during the 8 week study: Days 7, 14, 28
and 56 (± 1 day) and at a fifth time point if tapering-off is required. Subjects will be advised to contact the
investigator or site staff should their pain not be controlled and therefore may require additional in-clinic
visits. Telephone contacts will be made to subjects on Days 3, 6, and 9 to ensure adequate pain control is
achieved through dose titration and that possible side-effects are managed appropriately.

For eligible subjects, the following items will be recorded: standard demographic data; full medical,
surgical, and pain medication history; status of OA (including x-ray diagnostic of the diseased joint,
ACR Functional Class, OA classification per ACR criteria); the nature, dosage and evaluation of the
analgesic treatment of the past month. An x-ray diagnostic of OA taken within the last two years will be
acceptable.

Physical examinations will be recorded at the beginning and end of the study. Vital signs will be taken at
each visit. Height will be recorded at Visit 1. Weight will be recorded at Visit 1, 3, and 4. All adverse events
will be recorded from the first study-related procedure to the last study-related procedure. A statement that
the subject meets all eligibility criteria will be documented in the source notes by the Investigator.

The primary objective of this study is to determine the proportion of subjects who experience “moderate”,
“good”, or “excellent” pain control during 56 days of treatment with DRUG X. Secondary analyses will
include comparisons of the scores from the WOMAC Osteoarthritis Index questionnaire and SF-36 Acute
Health Survey (see Appendix B) during 56 days of treatment with DRUG X compared to baseline.
Physician and Subject Global Impression of Change scales (7-point scale, see Efficacy Evaluations
section) will be done at the Final Visit.

Efficacy Evaluations
Efficacy of DRUG X to treat the signs and symptoms of moderate to severe pain due to OA of the hip or
knee will be measured by:

*Pain Control Assessment.* Subjects will indicate the level of pain control at baseline and during the 14 day
treatment period with DRUG X at each visit and at each telephone contact. The question should be asked
at approximately the same time of day to ensure consistency. This consists of a five-point evaluation scale
from excellent to very poor. For this assessment the subject will be asked: “Think about the pain in your
________ (study joint). Would you rate your pain control today as being: excellent, good,
moderate, poor or very poor?”

*Functional Status.* Subjects will rate their pain, stiffness and physical function at baseline and each visit
by means of the WOMAC Questionnaire (Western Ontario and McMaster University Osteoarthritis
Index). A one-week recall period will be applied to all questions.
Quality of life. Subjects will complete a 36-item health survey used to evaluate the subject’s physical, social, mental, and general well-being at baseline and each visit by means of the SF-36 Acute Health Survey (see Appendix B). A one-week recall period will be applied to all questions.

Subject/Physician Global Impression of Change. At the completion of the study (or at the early withdrawal visit) the subject and the investigator will answer the question: “Since the start of the study, my [the subject’s] overall target joint status is?” - Very much improved, much improved, minimally improved, no change, minimally worse, much worse, very much worse.

Efficacy Criteria
The primary efficacy parameter of this study is the pain control of the target osteoarthritis hip or knee joint defined as a score of “excellent”, “good”, or “moderate” on the five-point scale: excellent, good, moderate, poor, and very poor. The proportion of subjects with pain control will be given per time point and at endpoint together with a 95% confidence interval. Outcomes are results from the WOMAC Osteoarthritis Index questionnaire, the SF-36 Health Survey and the Physician and Subject Global Impression of Change scores.

Safety Evaluations
All subjects will be considered for the safety evaluation. The incidence of all adverse events will be determined. Special attention will be given to those subjects who have discontinued the trial because of an adverse event, who experienced a severe or serious adverse event or who discontinued the trial due to lack of efficacy. Vital signs and the findings from physical examinations will be assessed.

The following will assess safety:
   a. Vital signs including sitting pulse and blood pressure (after a 5-minute rest), and respiratory rate will be measured at each visit.
   b. Weight will be recorded at Visit 1, 3, and 5 and height will be recorded at Visit 1 only.
   c. A complete medical history will be done at screening only and physical examinations will be done at screening and at Visit 4.
   d. Adverse events will be recorded from the time of the first study-related procedure to the time of the last study-related procedure.

II.2.L. Data analysis method
Efficacy Evaluations
Efficacy analyses will be carried out using the evaluable population. The evaluable population will consist of all subjects who have pain control information at Visit 2 (Week 2). In this case, missing values will be imputed using last observation carried forward (LOCF). Secondary analyses will also be carried out using the observed cases without imputation of missing values. Statistical tests will be carried out at the two-tailed 5% significance level unless specified otherwise.

The primary efficacy parameter will be the proportion of subjects on DRUG X achieving "excellent", "good" or "moderate" pain control compared to placebo on the 5 point scale: excellent, good, moderate, poor and very poor at Week 8 (Day 56). The results will be tabulated and plotted over time. Point estimates and 95% confidence intervals will be provided. A secondary analysis will be carried out using the observed cases.

Secondary analysis will be carried out using the evaluable population as well as the observed cases. Secondary responses include the WOMAC questionnaire, Acute SF-36 Health Survey Quality of Life questionnaire and Physician and Subject Global Impression of Change Scale. For the analyses using the evaluable population LOCF will be used to impute missing instrument scores. Tabulations will include
summary statistics such as the number of observed cases, the mean, standard deviation, minimum and maximum values.

WOMAC scores will be tabulated and plotted over time. Separate results will be tabulated for pain, stiffness, physical functioning and total scores. Significant differences between baseline and Week 8 will be assessed using the paired t-test. Acute SF-36 QoL scores will be tabulated and plotted over time. Significant differences between Baseline and Week 8 will be assessed using the paired t-test. Separate results will be tabulated for total score as well as for sub-scores for physical functioning, physical role limitation, emotional role limitation, social functioning, body pain, general mental health, vitality perception, and general health perception. Global Impression of Change scales provided by investigators and subjects will be tabulated.

Exploratory Analysis
The following tabulations and analysis will be presented as part of additional exploratory analyses.

a. A tabulation of the average and final titration doses.
b. A tabulation of the number of acetaminophen 500 mg tablets consumed per week.
c. A tabulation of the Treatment Assessment Questionnaire scores provided by subjects.
d. A comparison of the primary response between subjects taking NSAIDs and a weak opioid/acetaminophen combination and subjects taking only a weak opioid/acetaminophen combination prior to the study will be carried out using the exact Fisher test.

III. SUGGESTED READINGS

APPENDIX A. The American College of Rheumatology criteria for the classification and reporting of osteoarthritis

Classification Criteria for Osteoarthritis of the Hip

Traditional format

Hip pain plus at least two of the following:
- ESR of less than 20 mm per hour
- Femoral or acetabular osteophytes on radiographs
- Joint space narrowing on radiographs

Classification-tree format

Hip pain plus femoral or acetabular osteophytes on radiographs
or
hip pain plus joint space narrowing on radiographs and an ESR of less than 20 mm per hour

ESR = erythrocyte sedimentation rate


Classification Criteria for Idiopathic Osteoarthritis of the Knee

Traditional format

Knee pain plus osteophytes on radiographs and at least one of the following:
- Subject age older than 50 years
- Morning stiffness lasting 30 minutes or less
- Crepitus on motion

Classification-tree format

Knee pain and osteophytes on radiographs
or
knee pain plus subject age of 40 years or older, morning stiffness lasting 30 minutes or less and crepitus on motion.

ESR = erythrocyte sedimentation rate

APPENDIX B. Quality of Life (SF-36) Questionnaire

SF-36 ACUTE VERSION

INSTRUCTIONS: This survey asks for your views about your health. This information will help keep track of how you feel and how well you are able to do your usual activities.

Answer every question by marking the answer as indicated. If you are unsure about how to answer a question, please give the best answer you can.

1. In general, would you say your health is:

(circle one)

Excellent ..................................................................................................................1
Very good ...............................................................................................................2
Good ....................................................................................................................3
Fair ....................................................................................................................4
Poor ...................................................................................................................5

2. Compared with one week ago, how would you rate your health in general now?

(circle one)

Much better now than one week ago ...................................................................1
Somewhat better now than one week ago ............................................................2
About the same as one week ago ........................................................................2
Somewhat worse now than one week ago ...........................................................4
Much worse now than one week ago ...................................................................5

3. The following questions are about activities you might do during a typical day. Does your health now limit you in these activities? If so, how much?

(circle one number on each line)

<table>
<thead>
<tr>
<th>ACTIVITIES</th>
<th>Yes, Limited A Lot</th>
<th>Yes, Limited A Little</th>
<th>No, Not Limited At All</th>
</tr>
</thead>
<tbody>
<tr>
<td>a. Vigorous activities, such as running, lifting heavy objects, participating in strenuous sports</td>
<td>1</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>b. Moderate activities, such as moving a table, pushing a vacuum cleaner, bowling, or playing golf</td>
<td>1</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>c. Lifting or carrying groceries</td>
<td>1</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>d. Climbing several flights of stairs</td>
<td>1</td>
<td>2</td>
<td>3</td>
</tr>
</tbody>
</table>
e. Climbing one flight of stairs

f. Bending, kneeling, or stooping

g. Walking more than a mile

h. Walking half a mile

i. Walking one hundred yards

j. Bathing or dressing yourself

4. During the past week, have you had any of the following problems with your work or other regular daily activities as a result of your physical health?

(Circle one number on each line)

<table>
<thead>
<tr>
<th>YES</th>
<th>NO</th>
</tr>
</thead>
<tbody>
<tr>
<td>a. Cut down on the amount of time you spent on work or other activities</td>
<td>1</td>
</tr>
<tr>
<td>b. Accomplished less than you would like</td>
<td>1</td>
</tr>
<tr>
<td>c. Were limited in the kind of work or other activities</td>
<td>1</td>
</tr>
<tr>
<td>d. Had difficulty performing the work or other activities (for example, it took extra effort)</td>
<td>1</td>
</tr>
</tbody>
</table>

5. During the past week, have you had any of the following problems with your work or other regular daily activities as a result of any emotional problems (such as feeling depressed or anxious)?

(circle one number on each line)

<table>
<thead>
<tr>
<th>YES</th>
<th>NO</th>
</tr>
</thead>
<tbody>
<tr>
<td>a. Cut down on the amount of time you spent on work or other activities</td>
<td>1</td>
</tr>
<tr>
<td>b. Accomplished less than you would like</td>
<td>1</td>
</tr>
<tr>
<td>c. Didn't do work or other activities as carefully as usual</td>
<td>1</td>
</tr>
</tbody>
</table>

6. During the past week, to what extent has your physical health or emotional problems interfered with your normal social activities with family, friends, neighbours, or groups?

(circle one)

Not at all .................................................................................................................. 1
Slightly ......................................................................................................................... 2
Moderately .................................................................................................................... 3
Quite a bit ....................................................................................................................... 4
Extremely ..................................................................................................................... 5
7. How much **bodily** pain have you had during the **past week**?

(circle one)

None ................................................................................................................... ..........................1
Very mild.........................................................................................................................2
Mild ....................................................................................................................... ........................3
Moderate.................................................................................................................... ................4
Severe ..................................................................................................................... .........................5
Very severe................................................................................................................. .......................6

8. During the **past week**, how much did pain interfere with your normal work (including both work outside the home and housework)?

(circle one)

Not at all .................................................................................................................. ..................1
A little bit................................................................................................................ ...................2
Moderately.................................................................................................................. ...............3
Quite a bit ................................................................................................................. .................4
Extremely ................................................................................................................... ...............5

9. These questions are about how you feel and how things have been with you during the **past week**. For each question, please give the one answer that comes closest to the way you have been feeling. How much of the time during the **past week**:

(circle one number on each line)

<table>
<thead>
<tr>
<th></th>
<th>All of the Time</th>
<th>Most of the Time</th>
<th>A Good Bit of the Time</th>
<th>Some of the Time</th>
<th>A Little of the Time</th>
<th>None of the Time</th>
</tr>
</thead>
<tbody>
<tr>
<td>a. Did you feel full of life?</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>6</td>
</tr>
<tr>
<td>b. Have you been a very nervous person?</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>6</td>
</tr>
<tr>
<td>c. Have you felt so down in the dumps that nothing could cheer you up?</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>6</td>
</tr>
<tr>
<td>d. Have you felt calm and peaceful?</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>6</td>
</tr>
<tr>
<td>e. Did you have a lot of energy?</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>6</td>
</tr>
<tr>
<td>f. Have you felt downhearted and low?</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>6</td>
</tr>
<tr>
<td>g. Did you feel worn out?</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>6</td>
</tr>
<tr>
<td>h. Have you been a happy person?</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>6</td>
</tr>
<tr>
<td>i. Did you feel tired?</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>6</td>
</tr>
</tbody>
</table>
10. During the past week, how much of the time have your physical health or emotional problems interfered with your social activities (like visiting friends, relatives, etc.)? (Circle one)

All of the time...........................................................................................................................................1
Most of the time ...........................................................................................................................................2
Some of the time .........................................................................................................................................3
A little of the time ......................................................................................................................................4
None of the time.......................................................................................................................................5

11. How TRUE or FALSE is each of the following statements for you? (circle one number on each line)

<table>
<thead>
<tr>
<th></th>
<th>Definitely True</th>
<th>Mostly True</th>
<th>Don't Know</th>
<th>Mostly False</th>
<th>Definitely False</th>
</tr>
</thead>
<tbody>
<tr>
<td>a. I seem to get ill a little easier than other people</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
</tr>
<tr>
<td>b. I am as healthy as anybody I know</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
</tr>
<tr>
<td>c. I expect my health to get worse</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
</tr>
<tr>
<td>d. My health is excellent</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
</tr>
</tbody>
</table>
**APPENDIX C. Patient Diary**

1. Please record the number of tablets of EXTRA STRENGTH TYLENOL* you take each day for the pain in your ________________ (study joint). It is best if you can complete this information at the end of each day so you don’t forget to record any tablets you have taken.

   Please complete an entry for each day even if you did not require any EXTRA STRENGTH TYLENOL* and write down “0” for those days where you did not take any.

<table>
<thead>
<tr>
<th>Date DD-MON-YYYY</th>
<th>Number of tablets of Extra Strength Tylenol taken</th>
<th>Date DD-MMM-YYYY</th>
<th>Number of tablets of Extra Strength Tylenol taken</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Extra days, if applicable</td>
<td>Extra days, if applicable</td>
<td>Extra days, if applicable</td>
<td>Extra days, if applicable</td>
</tr>
</tbody>
</table>

*Tylenol is a register trademark of McNeil-PPC, Inc.*
Chapter 26. Osteoarthritis/arthrosis Long Term Studies: Delay in Structural Progression

Jean-Pierre Raynauld, M.D. ¹
Johanne Martel-Pelletier, Ph.D. ¹
Boulos Haraoui, M.D. ¹
Jean-Pierre Pelletier, M.D. ¹

¹ Osteoarthritis Research Unit
Centre hospitalier de l’Université de Montréal (CHUM)
Hôpital Notre-Dame
Département de Medicine
Université de Montréal
Montréal, Québec
CANADA
I. INTRODUCTORY REMARKS

Osteoarthritis (OA) is a disorder which can potentially affect all synovial joints. It is characterized by degeneration of articular cartilage and bone remodelling. The pathological changes can be focal or more generalized and these changes correlate poorly with clinical symptoms and signs. However, it has been suggested that asymptomatic OA, diagnosed radiologically, is a precursor of symptomatic disease. Osteoarthritis, particularly of large joints of the lower limbs is now widely recognized as a major cause of chronic disability in the population. Currently, there are inconsistencies in the classification of drugs for the treatment of OA and the indications for their use.

II. RATIONAL FOR STUDY DESIGN

II.1. Outline of a Typical Development Plan

Multi-center, randomized, double-blind, placebo-controlled, parallel group study in patients with mild to moderate primary knee osteoarthritis according to the American College of Rheumatology (ACR) criteria (Altman 1986, Appendix A), who have been completely withdrawn from their previous analgesics or anti-inflammatory medications or have been newly diagnosed with mild to moderate primary knee OA, and who are not currently taking any analgesics or anti-inflammatory medications. Alternatively, the hip joint may also be used as a target joint to evaluate OA progression. Hip OA is also defined according to specific ACR criteria (Altman 1991, Appendix A).

II.2. Long-Term Studies: Delay in Structural Progression

II.2.A. Objectives

The purpose of the study is to evaluate the efficacy and safety of continuous treatment of subjects with OA of the knee (or hip) over a two (2) year period with DRUG X versus reference DRUG Y (placebo, analgesic, NSAID) in reducing articular cartilage volume loss (or alternatively the joint space width loss) measured as a percentage change from Baseline.

The primary objective is to determine the efficacy and safety of the investigational DRUG X compared to the reference DRUG Y in patients with symptomatic knee (or hip) OA according to the ACR classification having pain more than one month.

II.2.B. Primary Endpoints

Knee is used as a signal joint to assess OA progression

The primary outcome measure is the percentage of cartilage volume loss in the medial compartment (femoral condyle and tibial plateau) from Baseline as assessed by MRI imaging. The method is described in detail in a previous publication (Raynauld 2003). Let the average % cartilage loss in DRUG X and comparator Y treated groups be denoted by X% and Y%, respectively. The parameter of interest is the mean difference of percentage loss between treatment groups relative to the mean percentage loss of the comparator treated group. This parameter is estimated by the ratio, r defined by:

\[ r = \frac{Y\% - X\%}{Y\%} \]

An independent reader will analyse each MRI image and will be blinded to the visit sequence of each image except for the Baseline image.
**Alternative primary endpoint for the knee as the signal joint**

An alternative endpoint would be to measure the knee minimal joint space width (minimal JSW) in mm on serial standardised radiographs. The weight-bearing posteroanterior film of both knees flexed at 30 degrees (Shuss view) proposed by Piperno et al (Piperno 1998) demonstrate high accuracy and reliability. For this specific view, the patellae of both knees must touch the film cassette, the toes pointed straight ahead vertically relative to the knee and the pelvis touch the table. The angle of knee flexion measured by a goniometer should approximate 30 degrees. With the aid of fluoroscopy, the x-ray beam is adjusted to obtain a horizontal tibial plateau. The interbone distance at the narrowest point (minimal JSW), the joint surface area (JSA), and the mean JSW can be automatically calculated using an image analysis computer and program developed by Cronozier (Cronozier 1995). For the minimal JSW, a loss of 0.17 mm at one year compared to Baseline may be expected.

**HIP is used as a signal joint to assess OA progression**

For the hip, a pelvic radiograph is obtained annually from patients with a weight-bearing position and standing at 1 meter from the x-ray source with a 20 degree internal foot rotation. Again, the interbone distance at the narrowest point (minimal JSW), the joint surface area (JSA), and the mean JSW can be automatically calculated using the same image analysis computer and program developed by Cronozier (Cronozier 1995).

**II.2.C. Secondary Endpoints**

**Knee is used as the signal joint**

Data obtained from the MR acquisitions at 6 months and 1 year will also be analysed in addition to the Baseline and 2 year data to derive a linear rate of the internal compartment cartilage volume change (degradation) over time. This linear change will be contrasted between the two treatment groups.

Besides the internal compartment, MRI volume measurements of the total cartilage, lateral compartment, medial femoral condyle and lateral femoral condyle will also be analysed.

**Other secondary efficacy parameters for both knee or hip OA studies:**

- Change in the Western Ontario McMaster (WOMAC) OA Index (Bellamy 1996) from Baseline to two years of either or all of the 3 following variables:
  - a) Pain subscale of the WOMAC
  - b) Functional Index of the WOMAC
  - c) Total WOMAC Index

- Quality of Life Questionnaire (SF-36 Acute) (see Chapter 25, Appendix B)
- Use of rescue acetaminophen with a Patient Diary (see Chapter 25, Appendix C)
- Patient and Investigator Global Assessment of Efficacy (Visual Analog Scale, 0 to 100 mm, 100 = worse).
- Number of subjects with an indication for knee or hip replacement surgery, as determined throughout the study.

**II.2.D. Study Design**

A multi-centre, randomised, double-blind, two-parallel group study comparing the comparator (placebo, analgesic, or NSAID) to DRUG X in subjects with OA of the knee or hip. Following a one week wash-out period, subjects will be treated with either DRUG X or comparator Y for 24 months.

At Visit 1 (Screening Visit) written informed consent will be obtained prior to any study-related evaluations or procedures. A physical examination, standing x-ray and laboratory tests will be performed and the inclusion and exclusion criteria will be reviewed. Females of childbearing potential (i.e. not post-menopausal or surgically sterilised) will have a urine pregnancy test. The result must be negative for the subject to continue in the study and the subject must agree to use an acceptable method of birth control for
the duration of the study. Eligible subjects will begin a one (1) week wash-out period during which all medications for OA (i.e. NSAIDS and analgesics, with the exception of acetaminophen which will be permitted as rescue medication) will be terminated.

At Visit 2 (Day 0) the inclusion and exclusion criteria will again be reviewed and eligible subjects will be randomised to either DRUG X twice daily or comparator Y twice daily.

At Visit 3 (Month 1) and Visit 4 (Month 2), and at all subsequent visits, subjects will undergo measurement of vital signs, and be assessed for adverse events and changes in concomitant medication (including acetaminophen use). Laboratory tests will be performed at Visit 3 (Month 1), Visit 6 (Month 6), Visit 8 (Month 12), Visit 10 (Month 18) and Visit 12 (Month 24).

At Visit 5 (Month 3), and every three (3) months thereafter, clinical assessments and questionnaires (i.e. SF-36, WOMAC OA Index, Subjects and Investigators Global Assessment of Efficacy and Tolerability) will be completed in addition to measurement of vital signs and assessment for adverse events and concomitant medication use.

MRI and/or radiological assessments will be performed at Visit 2 (Baseline), Visit 8 (Month 12) and Visit 12 (Month 24) of the treatment period. An additional MRI scan will be performed at Visit 6 (Month 6). Measurements of structural parameters will all be carried out centrally by the same evaluators.

Subjects must not use oral or parenteral anticoagulants (with the exception of ASA at a maximum daily dose of 325 mg), oral or topical NSAIDs (other than study medication), immunosuppressive drugs, lithium carbonate, phenytoin, analgesic drugs including over the counter preparations (other than rescue medication provided by the Investigator), other anti-arthritic drugs, including indomethacin, or compounds containing non-approved agents for arthritis or agents claiming to possess disease/structure-modifying properties (e.g. glucosamine and/or chondroitin sulfate containing compounds).

Acetaminophen, provided by the Sponsor, will be the sole analgesic medication permitted during the study. Acetaminophen will be allowed up to a maximum of 4 g daily, and will have to be stopped 24 hours before each study visit. The consumption of acetaminophen, a secondary parameter of efficacy, will be documented in the CRF. Subjects who require more than 4 g daily of acetaminophen must be discontinued from the study.

The medications and other treatments in use for intercurrent illnesses at the time of the Baseline Visit should remain constant for the duration of the study, as evaluated by the Investigator. A subject who is on an established physiotherapy regimen should continue with the same regimen during the study period.

II.2.E. Planned Sample Size for Knee MRI Changes Used as Primary Endpoint

We require 80% power to detect the clinically important difference of r (see primary outcome) as defined in the primary endpoint greater than or equal to 0.3 (30%). The required sample size at the two-sided 0.05 significance level is 110 per group. Assuming an approximate 25% dropout rate, the total sample size is 276 subjects (138 subjects per group).

This sample size estimation is based on the assumption that the comparator Y group will experience a 7.6% reduction of cartilage volume loss of the medial compartment after two years with a standard deviation of 6.0. The DRUG X treated subjects are expected to have a 5.32% reduction of cartilage volume loss after two years with a standard deviation of 6.0.
II.2.F. Study Population
Approximately 280 patients with moderate to severe knee OA will be randomised in this study. A total of 220 patients will be expected to complete the study.

II.2.G. Specific Inclusion Criteria
a. Ambulatory outpatients of either sex between 40 and 80 years of age inclusive and with primary OA of the knee who will not require surgical treatment for at least two years after inclusion.

b. Subjects with OA of the knee meeting the American College of Rheumatology (ACR) classification (see Chapter 25, Appendix A).

c. Subjects complaining of intermittent or constant pain for at least 50% of the time within two months prior to the Baseline Visit (Visit 2) and for whom treatment with NSAIDs is indicated.

d. Subjects with a WOMAC pain subscale index of at least 40 after a 24-hour washout of any analgesics and a seven (7) day washout of any NSAIDs.

e. Subjects with Kellgren and Lawrence Grade 2 or 3 OA of the knee (Kellgren 1957) assessed using a weight-bearing view radiograph taken not more than six (6) months prior to the Baseline Visit (Visit 2). Osteophytes may be present at either the medial or lateral, tibial or femoral margins.

f. The subject must have at least 2 mm and not more than 4 mm of medial joint space width, measured at the narrowest point in the medial compartment, as assessed locally by the Investigator with a caliper.

g. Subjects with at least one of the following three risk factors for increased risk of radiographic progression:
   • Body Mass Index (BMI) > 30
   • Heberden’s Nodes
   • Female gender

h. Subjects capable and willing to give written informed consent prior to enrolment, and who are able to understand and complete the study questionnaires.

i. Female subjects of childbearing potential (i.e. not post-menopausal or surgically sterilised) must have a negative urine pregnancy test at screening and must agree to use an acceptable method of birth control for the duration of the study.

II.2.H. Specific Exclusion Criteria
a. Subjects who have undergone total knee replacement in the contralateral knee within 6 months prior to the Screening Visit (Visit 1).

b. Subjects who have received an intraarticular corticosteroid injection in a lower joint during the three (3) months prior to the Baseline Visit (Visit 2) or any other injection (e.g. hyaluronic acid) within 90 days of the Baseline Visit (Visit 2).

c. Subjects with isolated lateral compartment disease defined by joint space loss in the lateral compartment only.

d. Subjects with OA secondary to a known disorder: Rheumatoid arthritis, seronegative spondylarthropathy, mixed connective tissue disease, collagen vascular disease, psoriasis, inflammatory bowel disease, recently clinically active (within three (3) months) CPPD or crystal-induced arthropathy (e.g. gout), any history of fracture involving the study joint, or any other type of arthritis.

e. Subjects with Class IV functional capacity using the American Rheumatism Association criteria.

f. Subjects who have had surgery in any lower limb joint within 365 days of the Baseline Visit (Visit 2) or arthroscopy, aspiration or lavage in any lower limb joint within 180 days of the Baseline Visit (Visit 2).

g. Subjects who have used indomethacin, or compounds containing non-approved agents for arthritis or agents claiming to possess disease/structure-modifying properties (other than glucosamine and/or chondroitin sulfate containing compounds), in the 14 days prior to the Baseline Visit (Visit 2).
h. Subjects who have used glucosamine and/or chondroitin sulfate containing compounds within:
   • seven (7) days prior to the Baseline Visit (Visit 2) if they have been taking the substance for
     one (1) month or less;
   • fourteen (14) days prior to the Baseline Visit (Visit 2) if they have been taking the substance
     for more than one (1) month but less than three (3) months;
   • ninety (90) days prior to the Baseline Visit (Visit 2) if they have been taking the substance
     for more than three (3) months
i. Subjects who have used medications with MMP-inhibitory properties (e.g. tetracycline or
   structurally related compounds) within 28 days prior to the Baseline Visit (Visit 2), or took
corticosteroids (systemic, >10 days duration) within 28 days of Visit 2.
j. Subjects who require acetaminophen at daily doses >4000 mg (4 g).
k. Subjects who are taking lithium carbonate, phenytoin or anticoagulants (with the exception of
   ASA up to a maximum daily dose of 325 mg).
l. Subjects who have received chondrocyte transplants in any lower extremity joint.
m. Subjects with comorbid conditions that restrict knee function.
n. Subjects with any significant diseases or conditions, including emotional or psychiatric disorders
   and substance abuse that, in the opinion of the Investigator, are likely to alter the course of OA or
   the subject’s ability to complete the study.
o. Subjects with any active acute or chronic infection requiring antimicrobial therapy, or serious
   viral (e.g. hepatitis, herpes zoster, HIV positivity) or fungal infections.
p. Subjects with a history of a gastrointestinal disorder that could prevent the subject from receiving
   NSAIDs for the total (2 years) duration of the study. Subjects may, at the Investigator’s
discretion, take proton pump inhibitors as required.
q. Subjects with pre-existing malignancy, other than basal cell carcinoma, within ten (10) years of
   the Screening Visit (Visit 1).
r. Subjects with chronic liver or kidney disease, as defined by AST or ALT 2 times the upper limit
   of normal (ULN), or by serum creatinine 2.0 mg/dL, at the Screening Visit (Visit 1) and at one
   repeat testing.
s. Subjects with a known hypersensitivity to DRUG X, NSAIDS (including NSAID comparator) or
   acetaminophen, or known acetaminophen- or NSAID-induced asthma.
t. Subjects receiving any investigational drug within 30 days prior to the Baseline Visit (Visit 2).
u. Subjects with any contraindications for undergoing MRI.

II.2.1. Tools for Assessing Endpoints

Primary Structural Endpoints
a. MRI
b. Standardised Radiograph (Schuss view)

Tools for assessing Secondary Endpoints
a. WOMAC and SF-36 Questionnaires,
b. Patient and Physician Global Disease Assessment,
c. Patient Diary.

Other Endpoints
Blood and urine samples will be collected at the Screening Visit, Visit 6 (Month 6), Visit 8 (Month 12),
Visit 10 (Month 18) and Visit 12 (Month 24). The following laboratory tests will be performed:

a. Haematology: haemoglobin, haematocrit, white blood cell count (WBC) (total and differential),
   red blood cell count (RBC), platelet count and erythrocyte sedimentation rate (ESR).
b. Chemistry: creatinine, urea, aspartate aminotransferase (SGOT/AST), alanine aminotransferase
   (SGPT/ALT), gamma-glutamyltransferase (gamma-GT), alkaline phosphatase, total bilirubin,
sodium, potassium and uric acid.
c. Urinalysis: pH, protein, glucose, ketone bodies, leukocytes, nitrite, haemoglobin/erythrocytes, urobilinogen and bilirubin. If haemoglobin/erythrocytes or protein are found in the urine, a microscopic investigation will be performed.

d. Urine pregnancy test will be performed on female subjects of child bearing potential at Visit 1 only. The urine pregnancy test must be negative in order for females of child bearing potential to proceed in the study. Females are considered of child bearing potential unless they are post-menopausal or surgically sterilised.

Frozen sera and urine will be collected at Visit 2 (Month 1), Visit 6 (Month 6), Visit 8 (Month 12), Visit 10 (Month 18) and Visit 12 (Month 24). Frozen samples will be stored until the end of the study and batch-shipped to the laboratory designated by the Sponsor. Ideally, samples will be stored at -70 degrees C.

All laboratory tests will be performed by the central laboratory. A copy of the laboratory report will be collected by the Sponsor.

Safety Considerations
All adverse events (AEs), whether observed by the Investigator or reported by the subject, will be recorded in the CRF provided by the Sponsor. The date of onset, duration, intensity, action taken due to the event, outcome, and relationship to the study medication will be indicated.

The severity of the event will be classified according to the following terms:
- a. mild: symptom is manifest but is tolerated
- b. moderate: normal activity affected
- c. severe: severe effect or inability to work or necessary to discontinue the study medication

The causality will be classified as:
- a. probable
- b. possible
- c. no causal relationship
- d. unclassified

For all AEs, the Investigator must obtain all information sufficient to determine whether the event meets the criteria for SAE reporting, and to determine the causality and outcome of the AE. Follow-up of all AEs reported during the study is required until the event resolves or stabilises at a level considered as acceptable by the Investigator and the Sponsor.

II.2.J. Specific Criteria for early Withdrawal and Discontinuation
Subjects may withdraw from the study at any time and for any reason. The Investigator may choose to withdraw a subject from the study if, in the opinion of the Investigator, continued participation in the study may compromise the safety or well-being of the subject. In the event of premature discontinuation from the study, the Investigator should determine the primary reason for discontinuation. A final evaluation should be completed at the time of discontinuation.

In the event that a subject discontinues the study at any time after randomisation, the Investigator will determine the reason for the discontinuation. Subjects who discontinue study medication will proceed to the Visit 12/End of Treatment visit and all assessments required for that visit will be performed.

If the subject discontinues from the study prior to Month 8 (Visit 12) but is willing to continue being followed in the study up to Month 12, he/she will undergo all assessments required at Visit 12/End of Treatment except the MRI and Schuss view knee radiographs unless these would have been done at the
subject’s next scheduled study visit. Subjects will be treated for OA at the Investigator’s discretion and will undergo the following procedures at Visit 8 (Month 12) only:
   a. MRIs
   b. Schuss view knee radiographs.
   c. Vital signs (including weight, blood pressure and pulse measured after five (5) minutes at rest in the seated position).
   d. If the subject is unwilling to continue in this manner or, in the opinion of the Investigator, it is contrary to the well-being of the subject to continue, then a final study visit will be done at which all assessments required at the Visit 12/End of Treatment Visit will be completed.

Subjects who withdraw more than one month after randomisation will not be replaced.

II.2.J. Data Analysis Method

Intent-To-Treat Subjects

A subject who is randomised to treatment and takes at least one dose of study medication will be included in the Intent-To-Treat population.

Per Protocol Subjects

A subject who completes the trial will be included in the Per Protocol population.

Analysis of Efficacy

The primary objective of this study is to evaluate the efficacy and safety of continuous treatment over a two (2) year period with DRUG X versus comparator Y in reducing articular cartilage volume loss measured as a percentage change from Baseline.

The primary efficacy assessment is based on the ITT sample size of all 276 subjects. Assume 80% of subjects in each group complete the trial (with expected % medial compartment cartilage losses equal to 7.6% and 5.32% in the comparator Y and DRUG X groups, respectively, at two years). Also assume 20% of subjects per group average only 1 year of study and are lost to follow-up (with % cartilage losses averaging 3.91% and 2.74% in the comparator Y and DRUG X groups, respectively). Then 134 subjects per group would be required to be followed, to achieve the overall 0.05 significance level with 80% power. Hence, the total sample size of 276 subjects is sufficient to meet the primary study objectives using the ITT group.

Baseline Values and Demographics

Characteristics of the subjects such as age, gender and race in each treatment group will be compared (including comparisons by centre), using the appropriate parametric tests. Efficacy variables assessed at Visit 2 are considered as the Baseline values and will also be compared.

Stratification is not used for the primary efficacy analysis. The design uses a completely randomized assignment plan. However, secondary analyses will be performed adjusting for influential covariates (using e.g.: the Analyses of Covariance) should they be out of balance at Baseline. The ANACOVA may also be used, secondarily, to investigate site differences.

Analytical Plan

Analysis of Primary Efficacy Outcome:

The statistic, \( S = \ln(1-r) \), where \( r \) is defined in the primary outcome section, will be used to test the primary hypothesis, which can be stated as follows:

Hypothesis: The expected value of \( S \) will be less than or equal to \( \ln(0.7) \).
Statistical Method: The comparison of treatment groups will use the normal approximation for the distribution of S. The comparison is based on the standardised value of S.

Descriptive Efficacy Analyses:
Standard analyses will include tabulations of means, percentages, standard deviations and confidence intervals. Data listings, and plots, will also be presented when useful. Analyses by visit will compare the two treatment groups on both a Per Protocol and an Intent-To-Treat basis.

The average of continuous variables, such as the percentage of cartilage volume loss, rescue use of acetaminophen, WOMAC OA index and all VAS scores will be analysed using standard normal theory.

Nominal variables, such as the number of subjects with an indication for knee replacement surgery will be analysed using Yates’ continuity corrected chi-square test.

**Missing Value Strategies**
To assess the sensitivity of the results to the presence of missing data, any missing scores will be imputed using two strategies. The last observation carried forward and at a given time point, the last available observation for the subject plus the average change from the time of last observation for the subject computed using subjects in the same treatment group who have non-missing observations.

**Analysis of Safety**
Randomised subjects receiving at least one dose of study medication will be included in the safety analyses. Safety analyses will be performed in terms of incidence and severity of adverse and/or unexpected events. These will be tabulated and compared between treatment groups. In addition, a complete listing of all reports of adverse and/or unexpected events will be presented.

Sitting blood pressure and heart rate will be summarised at each visit. Means, standard deviations, minimum and maximum individual values will be presented and compared.

Newly occurred abnormalities in the laboratory values and ECG results will be tabulated for each group. The proportion of subjects reporting such experiences will be compared.

**Safety Monitoring Committee**
An independent safety monitoring committee, not otherwise involved in the conduct of the trial, will be the primary data and safety advisory group for the Sponsor. The safety monitoring committee will review the results of the interim analysis and will periodically review study results, evaluate the treatments for excess adverse effects, determine whether the basic trial assumptions remain valid, and judge whether the overall integrity and conduct of the trial remain acceptable.

**III. EXAMPLES OF LANDMARK WELL DESIGNED TRIALS**


IV. SUGGESTED READINGS


Chapter 27. Rheumatoid Arthritis

Jean-Pierre Raynauld, M.D. ¹

Johanne Martel-Pelletier, Ph.D. ¹

Boulos Haraoui, M.D. ¹

Jean-Pierre Pelletier, M.D. ¹

¹ Osteoarthritis Research Unit
Centre hospitalier de l’Université de Montréal (CHUM)
Hôpital Notre-Dame
Département de Medicine
Université de Montréal
Montréal, Québec
CANADA
I. INTRODUCTORY REMARKS

Rheumatoid arthritis (RA) is a systemic inflammatory disease characterised by a proliferating synovitis leading to cartilage and bone destruction, and resulting in joint deformities and increasing functional limitations. Rheumatoid arthritis affects about 1% of the adult population. It occurs 2 to 3 times more frequently in women than in men. A great proportion of RA patients will rapidly develop major disabilities and almost 50% will experience work loss within ten years of diagnosis (1). Progression of the disease may also lead to premature death (2).

II. STUDIES FOR REGISTRATION OF NEW DRUGS

II.1. Outline of a typical plan

Evidence is ample that joint damage is an early phenomenon and progresses relentlessly over the years. Moreover, there is a direct causal link between the synovitis, the anatomical damage and disability (3). A general consensus has emerged that the key goals of therapy in RA are early rapid control of joint inflammation and prevention of joint destruction.

Therefore, a number of outcomes can be evaluated in clinical trials of RA (4-5):

a. Reduction in signs and symptoms
b. Improvement in functional disability
c. Prevention of anatomical radiographic damage
d. Safety issues
e. Considerations for short-term symptom modifying trials

Reduction in the signs and symptoms of RA

The goal is to demonstrate an improvement in signs of disease activity as well as symptoms.

Ordinarily, trials should be of a duration of at least six months, unless the product belongs to an already well-characterised pharmacological class (e.g., NSAIDs) for which trials of three-month periods are sufficient to establish efficacy for signs and symptoms. Six-month trials are desirable because RA is a disease of long duration. Interventions which provide only short-term, time-limited, benefit are unlikely to have overall value to patients.

In evaluating signs and symptoms, methods which evaluate response over time are preferable to methods which incorporate only the baseline value and the final observation, unless there is a reason to weigh last-visit symptoms rather than intermediary symptoms.

Several individual clinical or laboratory outcome measures can be used, but it has become the norm to use validated composite endpoints or indices of signs and symptoms. These composites may also be used to construct categorical endpoints for patient success or failure. For example, the Paulus criteria (6) or the more widely accepted American College of Rheumatology (ACR) definition of improvement (ACR 20, 50, 70) (7,8) are now the norm used to assess a patient’s response. ACR 20, 50 or 70 refers to an improvement of 20%, 50% or 70% in:

a. The number of swollen joints and
b. The number of tender joints and
c. 3 of the following measures:
d. Patient’s evaluation of pain on a visual analogue scale (VAS)
e. Patient’s global assessment on a VAS
f. Physician’s global assessment on a VAS
g. Health Assessment Questionnaire (HAQ) (see appendix 1)
h. Erythrocyte sedimentation rate (ESR) or C-Reactive Protein (CRP)

Example: Success for each patient in a six-month trial could be predefined as meeting the criteria for improvement over baseline in at least four of six monthly observations and not dropping out because of toxicity.

The use of the 66- or 28-joint count is appropriate for the evaluation of the swollen and tender joints (9).

A Major Clinical Response can also be defined as a continuous six-month period of success by the “ACR 70”. Therefore the trial duration should last a minimum of seven months for an agent expected to have a rapid onset of action, and longer for agents with less prompt effects.

A Complete Clinical Response or Remission refers to a Major Clinical Response coupled to radiographic arrest based on validated X-ray scoring systems (10,11). Complete clinical response connotes a benefit requiring ongoing drug therapy while remission is defined by the same result while off all antirheumatic drugs. The 1981 ACR remission criteria (12) require at least five of the following: morning stiffness less than 15 minutes, no fatigue, no joint pain by history, no joint tenderness or pain on motion, no swelling of joints or tendon sheaths, and erythrocyte sedimentation rate (ESR) less than 20 for males or less than 30 for females. The duration of the trial should be of at least one year or longer, depending on the rapidity of onset of the drug. Trials evaluating a complete clinical response should use a categorical endpoint (patient complete response or treatment failure) as the primary outcome measure.

Another validated composite score is the Disease Activity Score (DAS). The original DAS combines the Ritchie articular index, 44 counts of swollen joints, ESR and an assessment of the patient’s global health (13). The DAS28 is a validated subscale combining 28 swollen joint counts, 28 tender joint counts, ESR and the patient’s global health (14). The DAS28 is a continuous measure that was developed to assess the level of disease activity at a certain point in time (status score), which gives it an added value compared to the ACR response scores. However, its complicated formula requires the use of a computer or hand-held calculator, which are readily available. The DAS has also been validated against the ACR response criteria; therefore, these 2 validated measures can be used interchangeably when evaluating response (15,16).

Because most RA outcome measures have a high degree of subjectivity, the highest confidentiality inpatient and assessor blinding should be sought to achieve a credible inference. Therefore all clinical evaluations, including joint counts, should be assessed by an independent party with no knowledge of the subject’s history (blinded assessor).

Prevention of disability
Currently, the Health Assessment Questionnaire (HAQ) (17) (see appendix 1) and the Arthritis Impact Measure Scales (AIMS) (18) (http://www.qolid.org/public/aims/cadre/cadre.htm) are adequately validated measures for use as the primary outcome measure in assessing disability. Studies should be of a duration of at least six months, but interventions/drugs seeking this claim should be two to five years in duration and must concomitantly demonstrate improvement in signs and symptoms. Since the full effect of RA on a patient is not captured without the use of more general HR-QOL measures, a validated measure such as the SF-36 should also be collected and patients should not worsen on these measures over the duration of the trial (19, SF-36 Web page http://www.sf-36.com)(Appendix 2).

Prevention of structural damage
Prevention of structural damage is an important goal of RA therapy. Trials evaluating this outcome should be at least one year in duration and based on comparisons of films taken at one year (and subsequent yearly points) with those taken at baseline. All randomised patients should have films at both time points,
regardless of whether they are continuing treatment. Patients dropping out of the trial should have films taken at that time. Pre-specification of the handling of dropouts is especially important in these trials.

Different outcome measures can be used, as follows:

a. Slowing X-ray progression using the Larsen, the modified Sharp or another validated radiographic index (10,11);
b. Prevention of new X-ray erosions or maintaining an erosion-free state; and
c. Other measurement tools, such as MRI or ultrasonography, could be employed. However, because of the potential of the technique to identify small albeit statistically significant changes, the magnitude of the difference that would reflect actual patient benefit is unclear and needs to be established.

Because slowing of radiographic progression does not in itself define a direct patient benefit, it is expected that the agent would also demonstrate efficacy in one of the other claims (e.g., prevention of disability). However, some agents are not intended to affect acute inflammation but are designed to prevent or slow joint destruction by other means. Since the ultimate goals of slowing joint destruction are to improve symptoms and/or to preserve functional ability, slowing radiographic progression of disease is considered a surrogate marker for overall patient benefit in RA.

Safety issues
Every RA investigational therapy raises safety concerns. Whenever there is a potential for significant toxicity, long-lasting or delayed-onset, it is desirable to design the Phase 2 studies to provide a group of patients with longer term follow-up preceding the larger Phase 3 studies. Provisions for long-term follow-up can also be helpful in addressing issues of immunosuppression, opportunistic infections, neoplasia, and induction of autoimmune disease. Standard toxicity grading scales and stopping rules are also desirable in Phase 2.

Considerations for short-term symptom modifying trials
Phase 2 trials are used to better define the dose- and exposure-related activity and toxicity of the agent. Enough information should be generated to ensure that the Phase 3 trials are conducted safely and with a probability of success. Once a reasonably safe range of doses has been established, randomised, parallel-arm dose-comparison trials are usually recommended. The use of a placebo arm is desirable for several reasons. First, if no difference is found among doses, there is usually no other way to determine whether all doses were equally effective or equally ineffective. Second, if a dose-response trend is found, the placebo arm may indicate the possible magnitude of the observed effect. If use of a placebo is impossible, designs should include wide dose ranges or durations, or repetitions. Active controlled designs that specify an arm with a well-characterised known therapy can also be very useful.

The overall goal of Phase 3 work is to demonstrate the efficacy of the product in convincing controlled trials and to accrue a sufficient safety database. The new drug is tested against an accepted marketed drug in a randomised (frequently placebo-controlled) double blind trial.

II.2. Short-term trial in RA

II.2.A. Objectives
The primary objective is:

a. to evaluate the clinical efficacy of drug X versus placebo in the treatment of signs and symptoms of RA.

The secondary objectives are:

b. to evaluate the clinical efficacy of drug X versus drug Y based on pain intensity;
c. to determine an effective dose-range for drug X in the treatment of the signs and symptoms of RA, and enable dose selection for subsequent studies;
d. to evaluate the safety and tolerability of drug X in patients with RA;
e. to evaluate health outcomes data by using the EuroQol Questionnaire; and
f. to characterise the population pharmacokinetics of drug X in subjects with RA and assess the presence of a pharmacokinetic/pharmacodynamic relationship with clinical outcome.

II.2.B. Primary endpoints
a. The percentage of ACR20 responders

II.2.C. Secondary endpoints
a. Change from baseline in tender joint count (68 joints) at each scheduled visit;
b. Change from baseline in swollen joint count (66 joints) at each scheduled visit;
c. Change from baseline in physician’s global assessment of arthritis condition at each scheduled visit;
d. Change from baseline in patient’s global assessment of arthritis condition at each scheduled visit;
e. Change from baseline in functional disability index (HAQ) at each scheduled visit;
f. Change from baseline in CRP at each scheduled visit;
g. Percentage of patients discontinuing due to lack of efficacy; and
h. Average total daily dose of rescue medication.

II.2.D. Study design
The study will be a multicentre, randomised double-blind placebo and active controlled parallel group dose ranging and will evaluate the safety and efficacy of 3 oral doses of drug X (A mg, B mg and C mg) relative to placebo and drug Y (D mg) in subject with clinically active rheumatoid arthritis treated for 24 weeks.

All concomitant medications taken during the study will be recorded in the CRF with indication, dose, route of administration as well as the start date and the end date. Medication allowed and prohibited is specified below:

a. Rescue analgesic therapy: up to 3,000 mg of acetaminophen per day is allowed for pain relief. However, rescue medication is not allowed 12 hours prior to a clinic visit in order to minimise the effect of acetaminophen on study endpoints; and

b. Prohibited medications:
   • Gastroprotective agents: misoprostol, sulcralfate
   • Proton pump inhibitors
   • H2 blockers
   • Any analgesic or NSAID other than those allowed by the protocol
   • Oral corticosteroid (equivalent to > 10 mg of prednisone) or initiated within 4 weeks of baseline
   • Intra-articular injections

II.2.E. Planned sample
Sample size assumptions: the primary endpoint is the proportion of ACR 20 responders at 24 weeks. Assuming a drug X response rate of 45% and a placebo response rate of 25%, a sample size of XXX evaluable subjects per treatment group is sufficient to detect a 20% difference between the 2 groups with 90% power and a 5% significant level. Assuming a 20% dropout rate between randomisation and Week 24, a total of YYY subjects is required to be randomised in the study.

II.2.F. Study population
A sufficient number of male and female patients with ACR-defined RA will be screened for enrollment in order to randomise XXX subjects (XXX/5 per treatment group). The study will be conducted at centres in YYYY countries.

II.2.G. Inclusion criteria
a. Males and females of 18 years of age or older;
b. Fulfillment of the 1987 American College of Rheumatology (ACR) criteria for RA (Appendix E) with a disease duration ≥ 6 months;

c. Active disease defined by the presence of the following criteria (based on 66/68 joint counts):
   • 5 or more swollen joints at screening and baseline visit
   • 6 or more tender joints at screening and baseline visit

d. Subjects receiving oral corticosteroids must be receiving a stable dose equivalent to prednisone ≤ 10 mg/day for at least 12 weeks prior to screening;

e. Subjects who are currently receiving DMARD therapy (including methotrexate, sulfasalazine, hydroxychloroquine), must be on stable dose for at least 8 weeks prior to screening;

f. Women of child-bearing potential must use adequate contraceptive precautions (abstinence, oral contraceptives, barrier and spermicide, IUD or surgical sterilisation); and

g. Before any study-specific procedure, subjects must give informed consent for participation in the study.

II.2.H. Exclusion criteria

a. Known history of hypersensitivity or intolerance to NSAIDs, aspirin, COX-2 inhibitors or acetaminophen;

b. ACR functional class IV;

c. Receipt of any investigational drug within 28 days or five half-lives (whichever is longer) of study drug initiation;

d. Any clinical or laboratory abnormality found at screen or baseline which, in the opinion of the investigator, is clinically significant and would compromise the conduct or outcome of the study;

e. History of neoplasm or lymphoproliferative disease including lymphoma. Removed basal cell carcinoma is acceptable;

f. Active infection requiring antibiotic therapy;

g. Intra-articular, soft tissue, or intra-muscular corticosteroid injections during the 4 weeks prior to screening;

h. Initiation or change of dose of a standard DMARD within 12 weeks prior to baseline;

i. Use of oral corticosteroids at doses greater than the equivalent of 10 mg/day of prednisone or initiation of treatment within 4 weeks prior to baseline;

j. Initiation of, or change to, an established physiotherapy program within 2 weeks prior to baseline. An established program must continue unchanged throughout the study period;

k. Any disorder which compromises the ability of the subject to give written consent and/or to comply with study procedures; and

l. Any other condition which, in the investigator’s opinion, would preclude the subject from participating.

II.2.I. Tools for assessing endpoints

Analysis of primary efficacy endpoint

The primary efficacy endpoint is the percentage of ACR 20 responders (See Section II.1.a above, “Reduction in the signs and symptoms of RA”). Primary inferences will be based on the ITT population at Week 24, the last observation carried forward (LOCF) endpoint. Initially, an overall chi-square test will be performed to determine any drug X effect versus placebo. If this test is positive at the 5% significance level, the differences between each dose level of drug X will be investigated using logistic regression techniques adjusting for co-variates. In addition, dose response of drug X will be investigated.

Analysis of secondary efficacy endpoints

Each of the secondary variables will be analysed using the final model taken from the primary analysis. All secondary variables will be analysed as drug X versus placebo, and differences between drug X and drug Y will be investigated. Continuous secondary efficacy variables will be analysed using analysis of covariance. Results will be presented as adjusted means of 95% confidence intervals between each dose level of drug X and placebo. Binary efficacy endpoints will be analysed using logistic regression
techniques, adjusting for appropriate covariates. Results will be presented as adjusted odds ratios and 95% confidence intervals around the odds ratios.

**Safety**

Adverse events (AEs) will be coded using MedDRA. A summary of the number and percentage of subjects with the following adverse events will be displayed by treatment group: all AEs, drug-related AEs, serious AEs, and AEs leading to permanent discontinuation of the study drug.

Each laboratory value will be flagged to show whether it is within, below or above the normal range, and if it is of a clinical concern.

**II.2.J. Specific criteria for early withdrawal and discontinuation**

Randomised subjects who discontinue from the study prematurely will not be replaced. However, subjects who fail the screening period (screen failures) will be replaced.

If a subject is prematurely discontinued from participation in the study for any reason, the investigator must make every effort to perform the following procedures: completion of ACR20 assessments, vital signs, haematology and biochemistry laboratory assessments, and adverse events assessment.

**II.3. Long-term disease modifying trials**

Trials aiming to prevent structural radiographic damage should be of a duration of at least one year, with a second year extension to show sustainability of effect.

The study is designed as a one-year superiority trial to show efficacy of drug X in combination with MTX in patients with active RA and inadequate response to MTX, and maintained on MTX.

**II.3.A. Objectives**

The primary objective is:

- to compare the clinical efficacy of drug X used in combination with MTX versus MTX alone.

The secondary objectives are:

- structural damage progression;
- the proportion of patients with a Major Clinical response, and
- to evaluate the safety and tolerability of drug X in patients with RA.

**II.3.B. Primary endpoints**

- symptomatic relief as measured by the ACR20 response at six months;
- functional improvement as measured by the HAQ at one year, and
- structural radiographic damage as assessed by the total modified Sharp score at one year.

**II.3.C. Secondary endpoints**

- ACR 50 and 70 response at six months;
- ACR 20, 50 and 70 response at one year;
- Radiographic erosion and joint space narrowing score assessed by the modified Sharp method at one year;
- The proportion of patients with a Major Clinical response (ACR 70 for six continuous months);
- The disease activity measured by the DAS 28 at six months and one year;
- The functional improvement at six months measured by the HAQ, and
- The safety of drug X.
II.3.D. Study design
The study will be a multicenter, randomised double-blind placebo controlled parallel group of one-year duration to evaluate the safety and efficacy of drug X, relative to placebo in subjects with clinically active rheumatoid arthritis despite adequate background therapy with methotrexate. A second year extension is planned to gather further safety data.

Concomitant therapy:
  a. Subjects will continue their current MTX dose unchanged up to Month 6; during this period, only decreases due to toxicity will be allowed;
  b. All other DMARDs (with the exception of MTX) are not permitted during that period;
  c. NSAIDs including ASA are permitted, provided the dose is stable up to Month 6; decreases in NSAIDs dose are permitted only in case of toxicity;
  d. Oral corticosteroids (equivalent to > 10 mg of prednisone) are permitted provided the dose is kept stable during the first 6 months of study;
  e. Rescue analgesic therapy: acetaminophen – a combination of acetaminophen and narcotics may be used except 12 hours prior to a joint evaluation visit, and
  f. Intra-articular injections should be avoided; however, if necessary, up to 2 injections are permitted during the first 5 months. No injections are allowed between Months 5 and 6.

II.3.E. Planned sample
Sample size assumptions: The group receiving drug X in combination with MTX will be compared to the placebo control group receiving MTX alone. Sample size will be based on a 5% level (2 tailed) of significance. A total of XXX randomised in a 1:1 ratio will yield a 99% power to detect a 20% difference in ACR20 response between the two groups. Based on hierarchical testing procedure for the co-primary endpoints, this sample size will allow the detection of 18% difference in the HAQ response and a 60% difference in the total modified Sharp score with a power of 90%. These power calculations are based on the results observed in the Phase 2 trials.

II.3.F. Study population
A sufficient number of male and female patients with ACR-defined RA and fulfilling the Inclusion/Exclusion criteria will be screened for enrollment in order to randomise XXX subjects (XXX/2 per treatment group). The study will be conducted at centres in YYYY countries.

II.3.G. Specific inclusion criteria
  a. Males and females of 18 years of age or older;
  b. Fulfillment of the 1987 American College of Rheumatology (ACR) criteria for RA (Appendix E) with a disease duration ≥ 6 months;
  c. Active disease defined by the presence of the following criteria (based on 66/68 joint counts):
     • 10 or more swollen joints at screening and baseline visit
     • 12 or more tender joints at screening and baseline visit
     • CRP ≥ 1 mg/dL at screening
  d. Subjects receiving oral corticosteroids must be receiving a stable dose equivalent to prednisone ≤ 10 mg/day for at least 12 weeks prior to screening;
  e. Subjects must have been taking MTX for at least 3 months at a dose of ≥ 15 mg per week and be on a stable dose for the last 4 weeks;
  f. All other DMARDs with the exclusion of MTX must have been discontinued 4 weeks prior to baseline. In the case of infliximab, subject must have discontinued treatment at least 8 weeks prior to baseline;
  g. Women of child-bearing potential must use adequate contraceptive precautions (abstinence, oral contraceptives, barrier and spermicide, IUD or surgical sterilisation);
  h. Before any study-specific procedure, subjects must give informed consent for participation in the study.
II.3.H. Specific exclusion criteria

- a. Women who are pregnant or breast-feeding;
- b. Women of child-bearing potential unwilling or unable to use an acceptable method of birth control;
- c. ACR functional class IV;
- d. Receipt of any investigational drug within 28 days or 5 half-lives (whichever is longer) of study drug initiation;
- e. Any clinical or laboratory abnormality found at screen or baseline which, in the opinion of the investigator, is clinically significant and would compromise the conduct or the outcome of the study;
- f. History of neoplasm or lymphoproliferative disease including lymphoma; removed basal cell carcinoma is acceptable;
- g. Active bacterial infection requiring antibiotic therapy;
- h. Active TB requiring treatment within the previous 3 years; subjects with a positive PPD at screening need to have completed treatment for latent TB and to have a normal chest X-ray;
- i. Subjects with a known history of positivity for HIV, Hepatitis C and Hepatitis B; subjects with herpes zoster which resolved less than 2 months prior to enrolment;
- j. Intra-articular, soft tissue, or intra-muscular corticosteroid injections during the 4 weeks prior to screening;
- k. Use of oral corticosteroids at doses greater than the equivalent of 10 mg/day of prednisone or initiation of treatment within 4 weeks prior to baseline;
- l. Any disorder which compromises the ability of the subject to give written consent and/or to comply with study procedures, and
- m. Any other condition which, in the investigator’s opinion, would preclude the subject from participating.

II.3.I. Tools for assessing endpoints

Several individual clinical or laboratory outcome measures can be used to assess the changes in signs and symptoms, but it has become the norm to use validated composite endpoints or indices of signs and symptoms. These composites may also be used to construct categorical endpoints for patient success or failure. For example, the Paulus criteria (6) or the more widely accepted American College of Rheumatology (ACR) definition of improvement (ACR 20, 50, 70) (7,8) are now the norm used to assess a patient’s response. ACR 20, 50 or 70 refers to an improvement of 20%, 50% or 70% in:

- a. The number of swollen joints and
- b. The number of tender joints and
- c. 3 of the following measures:
- d. Patient’s evaluation of pain on a visual analogue scale (VAS)
- e. Patient’s global assessment on a VAS
- f. Physician’s global assessment on a VAS
- g. Health Assessment Questionnaire (HAQ) (see appendix 1)
- h. Erythrocyte sedimentation rate (ESR) or C-Reactive Protein (CRP)

Disability is assessed by using the Health Assessment Questionnaire (HAQ) (17) (see appendix 1) and the Arthritis Impact Measure Scales (AIMS) (18) (http://www.qolid.org/public/aims/cadre/cadre.htm). Since the full effect of RA on a patient is not captured without the use of more general HR-QOL measures, a validated measure such as the SF-36 should also be collected and patients should not worsen on these measures over the duration of the trial (19, SF-36 Web page http://www.sf-36.com)(Appendix 2).

The progression of structural damage is assessed by:

- a. Slowing X-ray progression using the Larsen, the modified Sharp or another validated radiographic index (10,11);
- b. Prevention of new X-ray erosions or maintaining an erosion-free state; and
c. Other measurement tools, such as MRI or ultrasonography, could be employed. However, because of the potential of the technique to identify small albeit statistically significant changes, the magnitude of the difference that would reflect actual patient benefit is unclear and needs to be established.

II.3.J. Specific criteria for early withdrawal and discontinuation
Randomised subjects who discontinue prematurely from the study will not be replaced. However, subjects who fail the screening period (screen failures) will be replaced.

Study therapy must be immediately discontinued for the following reasons:
   a. Withdrawal of consent;
   b. Any clinical or laboratory adverse event or intercurrent illness which, in the opinion of the investigator, indicates that continued treatment is not in the best interest of the patient;
   c. Inability of the patient to comply with the requirements of the protocol;
   d. Pregnancy; and
   e. Use of prohibited or restricted medications.

If, for any reason, a subject is prematurely discontinued from participation in the study, the investigator must make every effort to perform the following procedures: completion of ACR assessments, DAS evaluation, vital signs, haematology and biochemistry laboratory assessments, adverse events assessment and X-rays.

II.3.K. Data analysis method
Analysis of primary efficacy endpoints.
Co-primary analysis includes, in the order of sequential testing, comparisons between drug X and placebo in:
   a. ACR20 response at 6 months by chi-square test with correction for continuity. Percentage improvement in individual core components of the ACR response will be summarised by treatment;
   b. Functional improvement measured by the HAQ; a HAQ response is defined as a reduction of 0.3 unit from baseline; comparisons between treatment groups will be performed with a continuity chi-square test; and
   c. Structural radiographic damage progression measured by the total modified Sharp score; a non-parametric analysis of covariance model (ANCOVA) will be used to compare the changes from baseline of the scores.

Subjects who discontinue the study prematurely will be considered non-responders at all scheduled protocol visits subsequent to the point of discontinuation in the analysis of the ACR and the HAQ responses.

Analysis of secondary efficacy endpoints:
   a. The number and percentage of patients achieving a Major Clinical Response in each therapeutic group will be summarised for data collected up to Month 12 and the 95% confidence interval will be computed for the treatment difference; and
   b. Changes from baseline in the DAS28 score at Months 6 and 12 will be summarised by treatment arm, and the 95% confidence interval will be computed for the treatment difference.

Safety analysis
Adverse events (AEs) will be coded using MedDRA. A summary of the number and percentage of subjects with the following adverse events will be displayed by treatment group, as follows: all AEs, drug-related AEs, serious AEs, and AEs leading to permanent discontinuation of study drug.
Each laboratory value will be flagged to show whether it is within, below or above the normal range, and if it is of a clinical concern.

**III. SPECIAL CONSIDERATIONS FOR JUVENILE RHEUMATOID ARTHRITIS**

Juvenile Rheumatoid Arthritis (JRA) is a heterogeneous group of diseases that share the common feature of chronic, idiopathic synovitis, with onset prior to 16 years of age. These disorders have been divided into clinically distinct subsets based on the extent of joint involvement and extra-articular manifestations: pauci-articular, poly-articular, and systemic-onset JRA as well as oligoarthritis associated with HLA-B27, and they have been further subdivided based on clinical courses (20). Immunogenetic subsets appear to correlate with these clinical course subsets and are also distinct from adult RA (21). Of these various entities, polyarticular JRA is similar in many aspects to adult RA, particularly in clinical signs and symptoms, synovitis, and similar efficacy responses to some existing pharmacotherapy (NSAIDs, methotrexate, and prednisone). The application of principles in the conduct of clinical trials for adult RA largely applies to JRA as well, and this section outlines only those areas of difference from adult RA. The Committee on Drugs of the American Academy of Pediatrics has published guidelines for the ethical conduct of studies to evaluate drugs in paediatric populations (22) as well as general considerations for the clinical evaluation of drugs in infants and children (23).

As a general principle, children should not be subjected to an agent that has not been first tested for safety in adults. Testing may begin in children, but only when the anticipated benefits based on existing knowledge justify the anticipated risks. An agent developed specifically for use in JRA (e.g., a biological agent targeted against a specific pathogenic process that is unique to JRA and not present in adult RA) may need to be tested first in children as exposure in adult RA patients or even normal adult volunteers may be unrevealing.

**III.1. Reduction in the signs and symptoms of JRA**

All JRA trials should evaluate improvement based on a validated endpoint for improvement. Currently, the one validated approach is the definition of improvement established by the JRA core set: three of six (MD global, parent/patient global, number of active joints, number of joints with limited range of motion, functional ability, and ESR) improved by at least 30% and no more than one of six worsened by more than 30% (24). Because the JRA definition of improvement was validated using a trial of methotrexate, which primarily included polyarticular JRA patients, protocol individualisation may necessitate a refinement in the responder test for other patient subsets. For patients with systemic-onset JRA, additional assessment of fever, extra-articular manifestations, and thrombocytosis/leucocytosis may be useful coprimary endpoints (25). Outcome variables need to be appropriate and consistent with the type of agent under investigation. Investigators should specify, before the trial is initiated, the amount of change that is considered to be clinically important for each outcome variable. Trials should generally last at least six months, except when six-month efficacy data exist in adult RA and there are no reasons to expect loss of efficacy over time. Under these circumstances, trial durations may be blinded/randomised three-month periods; however, six-month periods of open safety data should be obtained. As with adult RA, a three-month trial duration is suggested for NSAIDs.

**III.2. Prevention of disability**

This claim is proposed to reflect durable improvement in physical function and disability in studies of one- to two-year periods with demonstrated improvement in signs and symptoms over the same period. Instruments currently validated for use in JRA include the Childhood Health Assessment Questionnaire (CHAQ), the Juvenile Arthritis Self-Report Index (JASI), and the Juvenile Arthritis Functional Assessment Report (JAFAR). HR-QOL should also be measured and shown not to worsen over the trial
duration. Endpoints should be tailored to subtypes enrolled in trials (e.g., to assess knee function in pauci-articular JRA patients in whom knee arthritis may be the primary arthritic manifestation). Instruments should be developmentally validated for the age ranges studied in a trial (26).

**IV. SUGGESTED READINGS**


**V. REFERENCES**

Appendix 1. Disability Index of the Health Assessment Questionnaire

Please check the one response which best describes your usual abilities OVER THE PAST WEEK:

<table>
<thead>
<tr>
<th>Without ANY Difficulty</th>
<th>With SOME Difficulty</th>
<th>With MUCH Difficulty</th>
<th>UNABLE To Do</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>DRESSING &amp; GROOMING</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Are you able to:</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>- Dress yourself, including tying shoelaces and doing buttons?</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>- Shampoo your hair?</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>ARISING</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Are you able to:</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>- Stand up from an armless straight chair?</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>- Get in and out of bed?</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>EATING</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Are you able to:</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>- Cut your meat?</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>- Lift a full cup or glass to your mouth?</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>- Open a new milk carton?</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>WALKING</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Are you able to:</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>- Walk outdoors on flat ground?</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>- Climb up five steps?</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Please check any AIDS or DEVICES that you usually use for any of these activities:

<table>
<thead>
<tr>
<th>Cane</th>
<th>Devices used for dressing (button hook, zipper pull, long-handed shoe horn, etc.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Walker</td>
<td>Built up or special utensils</td>
</tr>
<tr>
<td>Crutches</td>
<td>Built up or special chair</td>
</tr>
<tr>
<td>Wheelchair</td>
<td>Other (specify: _________________________)</td>
</tr>
</tbody>
</table>

Please check any categories for which you usually need HELP FROM ANOTHER PERSON:

<table>
<thead>
<tr>
<th>Dressing and Grooming</th>
<th>Eating</th>
</tr>
</thead>
<tbody>
<tr>
<td>Arising</td>
<td>Walking</td>
</tr>
</tbody>
</table>
Please check the one response which best describes your usual abilities **OVER THE PAST WEEK:**

<table>
<thead>
<tr>
<th></th>
<th>Without ANY Difficulty</th>
<th>With SOME Difficulty</th>
<th>With MUCH Difficulty</th>
<th>UNABLE To Do</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>HYGIENE</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Are you able to:</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>- Wash and dry your entire body?</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>- Take a tub bath?</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>- Get on and off the toilet?</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>REACH</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Are you able to:</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>- Reach and get down a 5 pound object (such as a bag of sugar) from just above your head?</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>- Bend down to pick up clothing from the floor?</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>GRIP</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Are you able to:</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>- Open car doors?</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>- Open jars which were previously opened?</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>- Turn faucets on and off?</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>ACTIVITIES</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Are you able to:</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>- Run errands and shop?</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>- Get in and out of a car?</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>- Do chores such as vacuuming or yard work?</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Please check any **AIDS** or **DEVICES** that you usually use for any of these activities:

<table>
<thead>
<tr>
<th>Device</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Raised Toilet Seat</td>
<td>Jar Opener (for jars previously opened)</td>
</tr>
<tr>
<td>Bathtub Seat</td>
<td>Long-Handled Appliances for Reach</td>
</tr>
<tr>
<td>Bathtub bar</td>
<td>Long-Handled Appliances in Bathroom</td>
</tr>
<tr>
<td>Other (specify: ________________________)</td>
<td>Other (specify: ________________________)</td>
</tr>
</tbody>
</table>

Please check any categories for which you usually need **HELP FROM ANOTHER PERSON:**

<table>
<thead>
<tr>
<th>Category</th>
<th>Assistance Needed</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hygiene</td>
<td>Gripping and Opening Things</td>
</tr>
<tr>
<td>Reach</td>
<td>Errands and Chores</td>
</tr>
</tbody>
</table>
Appendix 2: SF-36™ Health Status Survey

Patient Identification # __________________
Initials: ____/___/____
Date: ____/____/____

SHORT-FORM HEALTH SURVEY WORKSHEET

INSTRUCTIONS: This survey asks for your views about your health. This information will help keep track of how you feel and how well you are able to do your usual activities.

Answer every question by marking the answer as indicated. If you are unsure about how to answer a question, please give the best answer you can.

1. In general, would you say your health is:

   (circle one)
   
   Excellent ............................................................................................................................. 1
   Very good .......................................................................................................................... 2
   Good .................................................................................................................................. 3
   Fair ..................................................................................................................................... 4
   Poor ................................................................................................................................... 5

2. Compared to one year ago, how would you rate your health in general now?

   (circle one)
   
   Much better now than one year ago ................................................................................... 1
   Somewhat better now than one year ago ............................................................................ 2
   About the same as one year ago ....................................................................................... 3
   Somewhat worse now than one year ago .......................................................................... 4
   Much worse now than one year ago ................................................................................... 5

3. The following items are about activities you might do during a typical day. Does your health now limit you in these activities? If so, how much?

   (circle one number on each line)

<table>
<thead>
<tr>
<th>ACTIVITIES</th>
<th>Yes, Limited A Lot</th>
<th>Yes, Limited A Little</th>
<th>No, Not Limited At All</th>
</tr>
</thead>
<tbody>
<tr>
<td>a. Vigorous activities, such as running, lifting heavy objects, participating in strenuous sports</td>
<td>1</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>b. Moderate activities, such as moving a table, pushing a vacuum cleaner, bowling, or playing golf</td>
<td>1</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>c. Lifting or carrying groceries</td>
<td>1</td>
<td>2</td>
<td>3</td>
</tr>
</tbody>
</table>

333
4. During the past 4 weeks, have you had any of the following problems with your work or other regular daily activities as a result of your physical health?

(circle one number on each line)

<table>
<thead>
<tr>
<th></th>
<th>YES</th>
<th>NO</th>
</tr>
</thead>
<tbody>
<tr>
<td>a. Cut down on the amount of time you spent on work or other activities</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>b. Accomplished less than you would like</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>c. Were limited in the kind of work or other activities</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>d. Had difficulty performing the work or other activities (for example, it took extra effort)</td>
<td>1</td>
<td>2</td>
</tr>
</tbody>
</table>

5. During the past 4 weeks, have you had any of the following problems with your work or other regular daily activities as a result of any emotional problems (such as feeling depressed or anxious)?

(circle one number on each line)

<table>
<thead>
<tr>
<th></th>
<th>YES</th>
<th>NO</th>
</tr>
</thead>
<tbody>
<tr>
<td>a. Cut down on the amount of time you spent on work or other activities</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>b. Accomplished less than you would like</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>c. Didn’t do work or other activities as carefully as usual</td>
<td>1</td>
<td>2</td>
</tr>
</tbody>
</table>

6. During the past 4 weeks, to what extent has your physical health or emotional problems interfered with your normal social activities with family, friends, neighbors, or groups?

(circle one)

Not at all ................................................................................................................................. 1
Slightly ....................................................................................................................................... 2
Moderately ...................................................................................................................................... 3
Quite a bit ..................................................................................................................................... 4
Extremely ................................................................................................................................. 5
7. How much bodily pain have you had during the past 4 weeks?

(circle one)

None .......................................................... 1
Very mild .......................................................... 2
Mild .......................................................... 3
Moderate ............................................................. 4
Severe ............................................................. 5
Very severe .......................................................... 6

8. During the past 4 weeks how much did pain interfere with your normal work (including both work outside the home and housework)?

(circle one)

Not at all..................................................................................................................... 1
A little bit .................................................................................................................... 2
Moderately .................................................................................................................. 3
Quite a bit.................................................................................................................... 4
Extremely ................................................................................................................... 5

9. These questions are about how you feel and how things have been with you during the past 4 weeks. For each question, please give one answer that comes closest to the way you have been feeling. How much of the time during the past 4 weeks----

(circle one number on each line)

<table>
<thead>
<tr>
<th>ACTIVITIES</th>
<th>All of the Time</th>
<th>Most of the Time</th>
<th>A Good Bit of the Time</th>
<th>Some of the Time</th>
<th>A Little of the Time</th>
<th>None of the Time</th>
</tr>
</thead>
<tbody>
<tr>
<td>a. Did you feel full of pep?</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>6</td>
</tr>
<tr>
<td>b. Have you been a very nervous person?</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>6</td>
</tr>
<tr>
<td>c. Have you felt so down in the dumps that nothing could cheer you up?</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>6</td>
</tr>
<tr>
<td>d. Have you felt calm and peaceful?</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>6</td>
</tr>
<tr>
<td>e. Did you have a lot of energy?</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>6</td>
</tr>
<tr>
<td>f. Have you felt downhearted and blue?</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>6</td>
</tr>
<tr>
<td>g. Did you feel worn out?</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>6</td>
</tr>
<tr>
<td>h. Have you been a happy person?</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>6</td>
</tr>
<tr>
<td>i. Did you feel tired?</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>6</td>
</tr>
</tbody>
</table>
10. During the past 4 weeks how much did pain interfere with your normal work (including both work outside the home and housework)?

   (circle one)

   Not at all..................................................................................................................... 1
   A little bit....................................................................................................................... 2
   Moderately....................................................................................................................... 3
   Quite a bit......................................................................................................................... 4
   Extremely......................................................................................................................... 5

11. These questions are about how you feel and how things have been with you during the past 4 weeks. For each question, please give one answer that comes closest to the way you have been feeling. How much of the time during the past 4 weeks----

   (circle one number on each line)

12. During the past 4 weeks, how much of the time has your physical health or emotional problems interfered with your social activities (like visiting your friends, relatives, etc.)?

   (circle one)

   All of the time ................................................................................................................ 1
   Most of the time ............................................................................................................. 2
   Some of the time ............................................................................................................ 3
   A little of the time .......................................................................................................... 4
   None of the time ........................................................................................................... 5

13. How TRUE or FALSE is each of the following statements for you?

   (circle one number on each line)

<table>
<thead>
<tr>
<th>ACTIVITIES</th>
<th>Definitely True</th>
<th>Mostly True</th>
<th>Don't Know</th>
<th>Mostly False</th>
<th>Definitely False</th>
</tr>
</thead>
<tbody>
<tr>
<td>a. I seem to get sick a little easier than other people</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
</tr>
<tr>
<td>b. I am as healthy as anybody I know</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
</tr>
<tr>
<td>c. I expect my health to get worse</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
</tr>
<tr>
<td>d. My health is excellent</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
</tr>
</tbody>
</table>
Pharmacological Research in Other Disorders

Chapter 28. Neoplastic Diseases

Chapter 29. Analgesic Drugs for Cancer Pain Management

Chapter 30. Osteoporosis
Chapter 28. Neoplastic Diseases

Wendy Parulekar, M.D., FRCP(C) Med Onc

Clinical Trials Group
National Cancer Institute of Canada
10 Stuart Street
Queen's University
Kingston, Ontario K7L 3N6
CANADA
I. INTRODUCTORY REMARKS

Cancer is a global health problem. While there have been some major advances in the treatment of solid tumours over the past 30 years, most therapeutic successes have been relatively modest, leading to survival gains of a few percentage points at best. Thus there is an urgent need for continued clinical research involving new drugs or drug combinations. In acknowledgement of the intense resource and ethical implications of exposing patients with life threatening illnesses to experimental and potentially ineffective therapies, there exists an equally compelling need for well designed and conducted studies.

Chemotherapy is the mainstay of drug therapy for most solid tumours. These agents generally target DNA or the mitotic apparatus. In general, the principles behind chemotherapy dosing are based on preclinical data demonstrating that there is a direct relationship between dose and tumour cell kill as well as dose and toxicity. Thus selection of the appropriate dose for treatment represents a balance between antitumour effect and side effects.

The initial evaluation of new agents in human subjects occurs within the context of phase I and II studies. Phase I or dose finding studies seek to determine the appropriate dose for further study (recommended phase II dose or RPTD). This typically involves exposure of successive cohorts of 3 to 6 patients with various tumour types to increasing doses of drug. Careful evaluation of the toxicity and pharmacokinetic profiles of the new agent(s) occurs at each dose level. The RPTD is defined as that dose which produces serious but reversible side effects in a predefined proportion of patients. Subjects with advanced, heavily pretreated, disease are usually included in phase I trials provided that they have adequate organ and functional status as defined within the protocol.

Phase II studies screen for activity of a new drug or drug combination using the RPTD determined in the phase I study. Previously untreated or minimally treated patients with susceptible tumour types based on preclinical and early clinical evidence are included. The primary endpoint is estimation of the objective response rate, which is usually defined as the proportion of patients who have partial or complete shrinkage of tumour after drug exposure according to predefined standard criteria1 (1). In addition to characterization of the toxicity profile of the new therapy, these studies may also include pharmacokinetic or other pharmacodynamic endpoints. Many approaches exist for the conduct of phase II studies in cancer medicine including those that incorporate progression (2) and toxicity (3) information as part of the criteria for early termination of a trial. A common approach is that of the two stage design (4): if a minimum predetermined number of responses are seen with the first cohort of patients then the accrual is continued to the second stage to provide a more reliable estimation of activity. Sample size calculation is performed by specifying a response rate of interest as well as a lower response rate level below which the drug will be declared inactive. Traditional phase II design is non-comparative although randomization may be used to improve the efficiency of this type of study as a screening tool (5).

Phase III studies serve as the definitive tests of efficacy of new therapies. Using a randomized design to minimize bias, patients are allocated to the new agent(s) of interest or the standard therapy. In the field of oncology, these studies are powered to detect clinically meaningful differences in relevant endpoints such as overall or disease free survival. Quality of life or palliation of disease related symptoms may be the primary objective of symptom control studies. Phase III trials are generally resource intense due to the large sample sizes and duration of follow up generally required. Careful consideration must therefore be given to the primary study question and design since it is unethical to involve patients and investigators in trials addressing clinically irrelevant issues or involving poor methodology.

339
II. PHASE II STUDIES FOR REGISTRATION OF NEW THERAPIES

II.1. Outline of a typical development plan
Patients with advanced tumours of a single specific histological type are enrolled in a single arm, non-comparative study of a new drug/drug combination. The study may be conducted in a single or multi-institutional fashion although the latter is preferred to better estimate general feasibility of delivery of the new drug regimen.

II.2. Study plan

II. 2.A. Study Objectives
Primary objectives:
   a. To estimate the activity of drug X given in schedule Y (mg/route/frequency) in patients with previously untreated advanced tumours of a particular histology.

Secondary objectives:
   a. To assess the toxic effects (or “adverse effects”) of drug X in patients with previously untreated advanced tumours of a particular histology.
   b. In some studies consider also: To describe the relationship between molecular tumour characteristics and objective response.

II.2.B. Primary Endpoints
   a. Objective tumour response for solid tumours is assessed using the RECIST criteria. In phase II studies of new agents in hematological malignancies, response may be measured using peripheral blood indices (hemoglobin, white blood cell count, platelets, presence of malignant cells), bone marrow (cellularity and % of malignant cells) and cytogenetics.

II.2.C. Secondary Endpoints
   a. Duration of response.
   b. Adverse effects (toxic effects) in patients receiving drug X given in schedule Y as categorized and graded using the Common Terminology Criteria for Adverse Events (CTCAE) Version 3 (6).

II.2.D. Exploratory Endpoints
   a. Molecular measure of drug effect in tumour or surrogate tissue.

II.2.E. Study Design
Patients will be entered on this multi institutional, open label, single arm cohort study provided that all eligibility criteria are met and informed consent has been obtained.

Treatment must begin within 2 days of study entry. All patients will begin treatment at the protocol mandated dose and dose adjustments will be made on the basis of adverse (toxic) effects as required. Response and adverse event evaluation will be measured according to standard criteria as noted above.

All concomitant therapy, including alternative therapies, must be recorded on the case report forms. Other cytotoxic chemotherapy or investigational anti-cancer agents are not permitted.

II.2.F. Planned Sample Size
A typical sample size calculation will employ the Simon Two Stage Phase II Design method. Utilizing a response probability of interest (Ha) of 30%, a minimal response probability of 10% (Ho), error probabilities of 5% for accepting the drug with the minimal response probability and 20% for rejecting the drug with the response probability of interest:
Stage I of accrual: 10 response evaluable patients will be entered in the first stage. Using $H_0 \leq 10\%$ and $H_a \geq 30\%$, the drug will be declared inactive at the end of the first stage if there are fewer than 2 objective responses.

Stage II of accrual: If the above criterion is not met then 19 additional patients will be accrued onto the study for a final sample size of 29. The drug will be declared active if there are greater than 5 objective responses in the total sample.

II.2.G. Study Population
Previously untreated or minimally treated patients with susceptible tumour types based on pre clinical and early clinical evidence are included.

II.2.H. Specific inclusion criteria
Patients will be considered eligible for study entry provided that the following criteria are met:
  a. Histologically documented advanced/recurrent solid tumour of the specific histological type under evaluation.
  b. Presence of clinically or radiologically documented disease. At least one site of disease must be unidimensionally measurable defined as follows:
     • X-ray, physical exam $\geq 20$mm
     • Spiral CT scan $\geq 10$ mm
     • Non-spiral CT $\geq 20$ mm
  c. Patients must have a life expectancy of at least 12 weeks.
  d. Age $\geq 18$ years.
  e. Eastern Cooperative Oncology Group (ECOG) performance status of 0, 1, or 2 (See Appendix 1).
  f. Previous therapy: Prior adjuvant therapy is permitted but patients must not have had systemic therapy for advanced/recurrent disease.
     • Patients must be at least $\geq 6$ months since the last dose of adjuvant chemotherapy, if applicable.
     • Patients may have received prior radiation provided that all of the following conditions are met:
       i. There is measurable disease outside the previously irradiated area. Patients whose sole site of disease is in a previously irradiated area are ineligible unless there is evidence of progression or new lesions in the irradiated field.
       ii. At least 4 weeks must have lapsed since the last treatment with radiation.
       iii. Surgery is permissible provided that at least 4 weeks have lapsed since any major surgery.
  g. Laboratory Requirements: (must be within 7 days prior to study entry)
     • Hematology:
       i. Absolute granulocyte count (AGC) $\geq 1.5 \times 10^9$/L
       ii. Platelets $\leq 100 \times 10^9$/L
     • Chemistry:
       i. Serum creatinine $\leq$ Upper Normal Limit
       ii. Bilirubin $\leq$ Upper Normal Limit
       iii. AST $\leq 2.5 \times$ Upper Normal Limit
  h. Patient consent must be obtained according to local institutional policy of University Human Experimentation Committee requirements.
  i. Patients must be accessible for treatment and follow-up.

II.2.I. Specific Exclusion Criteria
  a. Prior history of other malignancies except: adequately treated non-melanoma skin cancer, curatively treated in-situ cancer of the cervix or other solid tumours curatively treated with no evidence of disease for $\geq 5$ years.
b. Prior chemotherapy for advanced/recurrent disease.
c. Non-measurable disease only.
d. Uncontrolled intercurrent illness including, but not limited to, ongoing or active infection, symptomatic congestive heart failure, unstable angina pectoris, cardiac arrhythmias or psychiatric illness/ social situation that would limit study compliance.
e. Concurrent treatment with anticancer or investigational agents including hormonal therapy.
f. Symptomatic brain metastases.
g. History of allergic reactions attributed to compounds of similar chemical or biologic composition to the study drug.
h. Pregnant or lactating women. All patients of child bearing potential must use adequate contraception while on study.

II.2.J. Tools for Assessment of Endpoints
Objective Response Criteria
Response and progression will be evaluated in this study using the RECIST criteria. Changes in only the largest diameter (unidimensional measurement) of the tumour lesions are used in the RECIST criteria.

Measurable Disease
Measurable lesions are defined as those that can be accurately measured in at least one dimension (longest diameter to be recorded) as ≥ 20 mm with conventional techniques (physical examination, CT, x-ray, MRI) or as ≥ 10 mm with spiral CT scan. All tumour measurements must be recorded in millimetres (or decimal fractions of centimetres).

Non-measurable Disease
All other lesions (or sites of disease), including small lesions (longest diameter < 20 mm with conventional techniques or < 10 mm with spiral CT scan) are considered non-measurable disease. Bone lesions, leptomeningeal disease, ascites, pleural/pericardial effusions, lymphangitis cutis/pulmonis, inflammatory breast disease, abdominal masses (not followed by CT or MRI) and cystic lesions are all non-measurable.

Target Lesions
All measurable lesions up to a maximum of 5 lesions per organ and 10 lesions in total representative of all involved organs should be identified as target lesions and be recorded and measured at baseline. Target lesions should be selected on the basis of their size (lesions with the longest diameter) and their suitability for accurate repetitive measurements (either by imaging techniques or clinically). A sum of the longest diameter (LD) for all target lesions will be calculated and reported as the baseline sum LD. The baseline sum LD will be used as reference to further characterize the objective tumour response of the measurable dimension of the disease. If there are > 10 measurable lesions, those not selected as target lesions will be considered together with non-measurable disease as non-target lesions.

Non-target Lesions
All non-measurable lesions (or sites of disease) plus any measurable lesions over and above the 10 listed as target lesions. Measurements are not required but these lesions should be noted at baseline and should be followed as “present” or “absent”.

All patients who receive at least one cycle of therapy and have their disease re-evaluated will be considered evaluable for response. All patients will have their BEST RESPONSE on study classified as outlined below:

Complete Response (CR): disappearance of all clinical and radiological evidence of tumour (both target and non-target).
Partial Response (PR): at least a 30% decrease in the sum of LD of target lesions, taking as reference the baseline sum LD.

Stable Disease (SD): steady state of disease. Neither sufficient shrinkage to qualify for PR nor sufficient increase to qualify for PD.

Progressive Disease (PD): at least a 20% increase in the sum of LD of measured lesions taking as references the smallest sum LD recorded since the treatment started. Appearance of new lesions will also constitute progressive disease. In exceptional circumstances, unequivocal progression of non-target lesions may be accepted as evidence of disease progression.

<table>
<thead>
<tr>
<th>Target Lesions</th>
<th>Non-Target Lesions</th>
<th>New Lesions</th>
<th>Overall Response</th>
<th>Best Response for this category also requires</th>
</tr>
</thead>
<tbody>
<tr>
<td>CR</td>
<td>CR</td>
<td>No</td>
<td>CR</td>
<td>≥ 4 wks. confirmation</td>
</tr>
<tr>
<td>CR</td>
<td>Non-CR/Non-PD</td>
<td>No</td>
<td>PR</td>
<td>≥ 4 wks. confirmation</td>
</tr>
<tr>
<td>PR</td>
<td>Non-PD</td>
<td>No</td>
<td>PR</td>
<td>documented at least once ≥ 4 wks. from baseline</td>
</tr>
<tr>
<td>SD</td>
<td>Non-PD</td>
<td>No</td>
<td>SD</td>
<td>no prior SD, PR or CR</td>
</tr>
<tr>
<td>PD</td>
<td>Any</td>
<td>Yes or No</td>
<td>PD</td>
<td></td>
</tr>
<tr>
<td>Any</td>
<td>PD*</td>
<td>Yes or No</td>
<td>PD</td>
<td></td>
</tr>
</tbody>
</table>

* In exceptional circumstances, unequivocal progression in non-target lesions may be accepted as disease progression.

Note: Patients with a global deterioration of health status requiring discontinuation of treatment without objective evidence of disease progression at that time should be reported as “symptomatic deterioration”. Every effort should be made to document the objective progression even after discontinuation of treatment.

Quality of life is measured using validated instruments such as the EORTC QLQ C30 (7). Patient diaries are appropriate for symptom control studies.

Specific Criteria for Early Withdrawal and Discontinuation: See Phase II Section.

Tools for Assessment of Primary Endpoints:
The same method of assessment and the same technique should be used to characterize each identified and reported lesion at baseline and during follow-up.

Clinical Lesions. Clinical lesions will only be considered measurable when they are superficial (e.g. skin nodules, palpable lymph nodes). For the case of skin lesions, documentation by colour photography including a ruler to estimate the size of the lesion is recommended.

Chest X-ray. Lesions on chest X-ray are acceptable as measurable lesions when they are clearly defined and surrounded by aerated lung. However, CT is preferable.

CT, MRI. CT and MRI might be the best currently available and reproducible methods to measure target lesions selected for response assessment. Conventional CT and MRI should be performed with cuts of 10 mm or less in slice thickness contiguously. Spiral CT should be performed using a 5 mm contiguous reconstruction algorithm. This applies to the chest, abdomen and pelvis. Head & neck and extremities usually require specific protocols.
Ultrasound. When the primary endpoint of the study is objective response evaluation, ultrasound (US) should not be used to measure tumour lesions that are clinically not easily accessible. It is a possible alternative to clinical measurements for superficial palpable nodes, subcutaneous lesions and thyroid nodules. US might also be useful to confirm the complete disappearance of superficial lesions usually assessed by clinical examination.

Endoscopy, Laparoscopy. The utilization of these techniques for objective tumour evaluation has not yet been fully and widely validated. Their uses in this specific context require sophisticated equipment and a high level of expertise that may only be available in some centres. Therefore, the utilization of such techniques for objective tumour response should be restricted to validation purposes in reference centres. However, such techniques can be useful to confirm complete pathological response when biopsies are obtained.

Tumour Markers. Tumour markers alone cannot be used to assess response. If markers are initially above the upper normal limit, they must normalize for a patient to be considered in complete clinical response. Specific additional criteria for standardized usage of PSA and CA-125 response in support of clinical trials are being developed.

Cytology, Histology. These techniques can be used to differentiate between PR and CR in rare cases (for example, residual lesions in tumour types such as germ cell tumours, where known residual benign tumours can remain).

The cytological confirmation of the neoplastic origin of any effusion that appears or worsens during treatment when the measurable tumour has met criteria for response or stable disease is mandatory to differentiate between response or stable disease (an effusion may be a side effect of the treatment) and progressive disease.

Tools for Assessment of Secondary Endpoints:
Response Duration. Response duration will be measured from the time measurement criteria for CR/PR (whichever is first recorded) are first met until the first date that recurrent or progressive disease is objectively documented.

Stable Disease Duration. Stable disease duration will be measured from the time of start of therapy until the criteria for progression are met, taking as reference the smallest measurements recorded since the treatment started.

Adverse Events. All patients will be evaluable for assessment of adverse events from the time of their first dose of study drug.

Adverse events will be monitored on an ongoing basis by the study principal investigator and, if applicable, the study coordinating office. Adverse events will be categorized using the CTCAE Version 3.0. The worst event for each patient in each category or subcategory will be described. Events related and unrelated to treatment will be captured.

II.2.K. Specific Criteria for Early Withdrawal and Discontinuation
Patients may stop protocol treatment in the following instances:
 a. Intercurrent illness which would, in the judgement of the investigator, affect assessments of clinical status to a significant degree and require discontinuation of protocol therapy.
 b. Unacceptable toxicity: Patients with intolerable or limiting toxicity despite dose reductions as defined in the protocol may be removed from study as well as those patients with Grade 3 or 4 toxicities which do not improve to ≤ Grade 1 despite drug hold for 2 weeks.
c. Tumour progression as defined by RECIST criteria.

d. Patient request.

II.2.L. Data Analysis Method

Enrollment will occur using a 2-stage design described above. Using the null hypothesis that the response rate is 10% and the alternate hypothesis that the response rate is 30%, the sample size will yield a significance level of 5% and a power of 80%.

Primary Endpoints:
Using the primary endpoint of overall survival in study of patients with advanced disease as an example, the data analysis will involve the generation of survival curves for each treatment arm. All randomized patients will be included in the primary analysis. Survival will be defined as the time from randomization to death from any cause. Patients who are alive at the time of final analysis will be censored at the time of last contact. A Kaplan-Meier curve for proportions of survival in each treatment arm will be displayed and 95% confidence intervals for median survival computed using the method of Brookmeyer and Crowley (8). The two treatment arms will be compared using the log-rank test adjusted for the stratification variables. In addition, the effect of study centre and other potential prognostic factors on overall survival will be assessed using Cox regression. The Schoenfeld residual plots will be used to check the model assumption for the Cox regression (9).

Secondary Endpoints:
Progression free survival (PFS):
This is defined as the time from randomization to the first observation of disease progression according to the RECIST criteria or death due to any cause. A patient who stops treatment with study drug and receives alternative therapy prior to documentation of disease progression will be censored on the date that the alternative therapy begins. If a patient has not progressed or received alternative therapy, PFS will be censored on the date of the last disease assessment. All analyses for overall survival will be similarly performed for PFS.

Response Rate:
Patients will be evaluable for objective tumour response if they have at least one measurable lesion at baseline and at least one disease assessment after baseline. In addition, patients who develop PD prior to this time will also be considered evaluable for response. The response rate will be estimated as the proportion of patients evaluable for response who meet the criteria for complete or partial response. A Cochran-Mantel-Haenszel test will be used to compare tumour response rate between arms adjusting for stratification factors.

The duration of response will be calculated for all patients achieving a PR or CR. Duration of PR/CR is defined as the time from first objective status assessment of CR/PR to the first time disease progression or death is documented. A patient who stops treatment with all study drugs and receives alternative therapy prior to documentation of disease progression, will be censored on the date alternative therapy begins. The date of progression will be considered as the event date for the duration of response. If a patient has not progressed or died, the duration of response will be censored on the date of the last known alive date. The duration of response will be analyzed using similar methods described for overall survival.

Quality of Life:
The quality life of patients will be assessed using EORTC QLQ-C30. The EORTC QLQ-C30 is a validated and reliable self-administered cancer specific questionnaire with multi-dimensional scales. It consists of both multi-item scales and single item measures, including five functioning domains, a global quality of life domain, three symptom domains and six single items. For each domain or single item measure a linear transformation will be applied to standardize the raw score to range between 0 and 100.
Since quality of life will be assessed longitudinally, the method of analysis of variance for repeated measures will be used for domains represented by aggregate scores.

Toxicity:
All patients who receive at least one dose of protocol therapy will be included in the safety analysis. Descriptive summary tables will be presented on safety parameters by treatment arm. Toxicity rates will be compared between treatment arms using the Fisher’s Exact Test, as needed. Oversight of the study by an independent data safety monitoring committee (DSMC) will occur. Included in the mandate of this committee will be ongoing review of the toxicity experience on trial and any interim analysis results as specified in the protocol (10).

III. PHASE III STUDIES FOR REGISTRATION OF NEW THERAPIES

III.1. Outline of a typical development plan

A promising drug/drug combination is selected for further study based on preclinical and early clinical evidence. Using an active and tolerable dose of therapy as defined in phase I and II studies, the relevant patient population (histological subtype, disease stage) and study question are chosen for the phase III trial.

III.2. Study plan

III.2.A. Study Objectives
As for Phase II studies.

III.2.B. Primary endpoints
Overall survival

III.2.C. Secondary endpoints
a. Disease free survival (adjuvant trials).
b. Progression free survival (advanced disease trials).
c. Toxicity.
d. Response rate and duration of response (advanced disease setting).
e. Quality of life.

III.2.D. Exploratory endpoints
a. Relationship between molecular characteristics of tumour and prognosis.
b. Relationship between molecular characteristics of tumour and probability of response to therapy.

III.2.E. Study Design
A randomized parallel group design is used. Blinding of treatment assignments may be appropriate, particularly in studies involving quality of life or symptom control endpoints.

Prior therapy may be allowed depending on the disease under evaluation but, in general, previously untreated or minimally treated patients with good functional status are assigned to the experimental or control arm as defined by the protocol.

Treatment: For advanced disease studies, therapy is generally continued until progression or occurrence of dose limiting toxicity. In early disease or adjuvant studies, protocol therapy is given for a fixed number of cycles or duration.
Treatment must begin within 2 days of study entry. All patients will begin treatment at the protocol mandated dose and dose adjustments will be made on the basis of adverse (toxic) effects as required. Response and adverse event evaluation will be measured according to standard criteria as noted above.

All concomitant therapy, including alternative therapies, must be recorded on the case report forms. Other cytotoxic chemotherapy or investigational anti-cancer agents are not permitted.

**III.2.F. Planned Sample Size**
Sample size calculations for clinical trials require specification of the type I and II errors as well as the magnitude of difference in outcome that the trial is designed to detect. The latter specification is a clinical one and depends on what difference in efficacy is likely to be present between the treatments and what difference would change current practice if detected.

As an example, a study in the advanced disease setting is designed to compare overall survival between patients randomized to ARM 1 (control) and patients randomized to ARM 2 (experimental). Based on other clinical data, the median survival of patients randomized to ARM 1 is estimated to be 0.55 years. In order to have 80% power to detect a 33% improvement in median survival in the experimental arm (hazard ratio of 1.33) using a two-sided 5% significance test, 381 deaths must be observed before the final analysis. In anticipation of accrual of 450 patients in 9 months, the required number of deaths (381) would be observed after following all patients for another 18 months.

Some studies employ interim analyses with adjusted p values to detect early and potentially clinically significant differences in outcome between arms before the final analysis. If interim analyses are planned, this should be clearly indicated in the body of the protocol.

Balance between the treatment arms for known prognostic factors is achieved by use of stratification factors at the time of randomization. Examples of typical stratification factors in oncology include disease stage, sex and performance status.

**III.2.G. Study Population**
Using a randomized design to minimize bias, patients are allocated to the new agent(s) of interest or the standard therapy.

**III.2.H. Specific inclusion criteria**
Inclusion Criteria: See Phase II section for typical eligibility criteria.

For advanced studies, prior systemic therapy may be allowed as long as an appropriate time lapse has occurred at the time of randomization. For adjuvant studies, no prior systemic therapy is generally allowed.

Patients with measurable or unmeasurable disease may be appropriate for advanced disease studies in which survival is the primary endpoint. Adjuvant study protocols for patients with early disease mandate the absence of tumour (clinically and radiologically) at the time of randomization.

**III.2.I. Specific Exclusion Criteria**
See Phase II section for typical ineligibility criteria

**III.2.J. Tools for Assessment of Endpoints**
Tools for Assessing Primary Endpoints:
For overall survival, the date of death is determined from hospital records or the death certificate whenever possible.
Tools for Assessing Secondary Endpoints:
In early disease studies measuring disease free survival, the date of relapse is the first date of clinical or radiological relapse.

Recurrence will be categorized as local, regional or distant, based on the histology and location of the primary tumour. The date of first recurrence should always be based on the onset of a sign of recurrence rather than the onset of a symptom.

The date of first detection of a palpable lesion is acceptable only when the diagnosis of tumour involvement is subsequently established.

The diagnosis of recurrent disease by radiographs or scans should be dated from the date of the first positive record, even if this is determined in retrospect.

For toxicity and response assessment, see Phase II Section regarding use of appropriate tools.

Quality of life is measured using validated instruments such as the EORTC QLQ C30 (7). Patient diaries are appropriate for symptom control studies.

III.2.K. Specific Criteria for Early Withdrawal and Discontinuation
See Phase II Section

III.2.L. Methods of Measurement
See Phase II Section

IV. EXAMPLES OF LANDMARK WELL DESIGNED TRIALS

IV.1. Phase II studies

IV.2. Phase III studies


V. SUGGESTED READINGS

V.1. Phase II studies


V.2. Phase III studies


VI. REFERENCES

APPENDIX I - PERFORMANCE STATUS (ECOG)

Grade

0  Fully active, able to carry on all pre-disease performance without restriction (Karnofsky 90-100).

1  Restricted in physically strenuous activity but ambulatory and able to carry out work of a light or sedentary nature (eg, light housework, office work) (Karnofsky 70-80).

2  Ambulatory and capable of all self-care but unable to carry out any work activities. Up and about more than 50% of waking hours (Karnofsky 50-60).

2  Capable of only limited self-care, confined to bed or chair more than 50% of waking hours (Karnofsky 30-40).

3  Completely disabled. Cannot carry on any self-care. Totally confined to bed or chair (Karnofsky 10-20).
Chapter 29. Analgesic Drugs for Cancer Pain Management

Kim Fisher, Ph.D. 1,2

Neil Hagen, M.D. 1,3

Patrick du Souich, M.D., Ph.D. 4

John Oldenhof, Ph.D. 5

Myroslava Romach, M.D. 5

Anh Ho Ngoc, Ph.D. 6

Edward Sellers, M.D., Ph.D. 5,7

1 Alberta Cancer Board Palliative Care Research Initiative
2 Department of Obstetrics and Gynecology
   University of Calgary
   Calgary, Alberta
   CANADA
3 Division of Palliative Medicine
   Department of Oncology, and Departments of Clinical Neurosciences and Medicine
   University of Calgary
   Calgary, Alberta
   CANADA
4 Département de pharmacologie
   Faculté de médecine
   Université de Montréal
   Montréal, Québec H3C 3J7
   CANADA
5 Ventana Clinical Research Corporation
   Toronto, Ontario
   CANADA
6 International Wex Technologies Inc.
   Vancouver, British Columbia
   CANADA
7 Departments of Pharmacology, Medicine and Psychiatry
   University of Toronto
   Toronto, Ontario
   CANADA
I. INTRODUCTORY REMARKS

Pain directly related to cancer or caused by treatments for cancer is a highly prevalent clinical problem. Between 30-85% of patients experience pain at some point during the illness trajectory, with estimates of 18-78% experiencing substantial pain in developed countries.

The current therapies for severe pain in cancer patients remain unsatisfactory. Opioids, the mainstay of cancer pain therapy, have a number of limitations. First, they are accompanied by a significant risk of dose limiting toxicity such as constipation, nausea and drowsiness. Second, some types of cancer pain, such as neuropathic pain and movement related bone pain, are difficult pain problems and carry a less good prognosis for control by opioids at doses that are tolerable. Third, opioids carry a significant perceived risk of abuse potential by society in general, which may at times represent barriers to the effective management of cancer pain. Fourth, cancer patients are at high risk for organ failure such as renal impairment or liver disease as their underlying disease progresses, and this comorbidity increases the risk of multifactorial, dose limiting symptoms associated with opioid administration such as delirium.

For cancer pain problems that have limited responsiveness to opioids, the World Health Organization (WHO) recommends the use of non-opioid (e.g. non-steroidal anti-inflammatory drugs) and adjuvant analgesics, in addition to opioids. Adjuvant analgesics include tricyclic antidepressants, anticonvulsants, N-methyl-D-aspartate antagonists, corticosteroids and others. The likelihood of effectiveness of many of the major classes of analgesics for difficult cancer pain problems remains disputed. There is an urgent need for controlled trials to evaluate the efficacy and adverse effect profiles of currently available non-opioid and adjuvant analgesics for cancer pain. There is also a need for newer analgesic drugs that have greater efficacy for severe cancer pain, better adverse effect profiles, and a reduced potential for toxicity, tolerance and addiction.

This chapter profiles a randomized, double-blind, placebo-controlled parallel-group trial design to evaluate the efficacy and safety of investigational drug XXX for patients with moderate to severe cancer pain. Based on the results from previous clinical studies, it is known that this drug has several important properties. Specifically, drug XXX has a rapid onset of action (5-30 minutes post injection), cumulative analgesic effect with twice-daily injections, long duration of action extending for days or weeks beyond the treatment period, and a favourable adverse effect profile at the studied doses. The trial described in this chapter is designed to better characterize these properties of drug XXX. For specific aspects of clinical trial design, the authors will also broaden the discussion to alternative approaches to help guide the development of Phase II clinical trials for analgesic drugs with different pharmacokinetic and pharmacodynamic properties. The examples in this chapter are most applicable to trials for evaluating analgesic drugs intended to be used in addition to patients’ current analgesic therapy.

II. PHASE II STUDIES FOR REGISTRATION OF NEW ANALGESIC DRUGS

II.1. Outline of a typical development plan

Early Phase II analgesic studies are typically multiple-dose clinical trials designed to identify the optimal dose for later Phase II studies, and to characterize the drug’s analgesic and adverse effect profiles. Once the optimal dose has been identified, later Phase II analgesic studies tend to be multicenter, randomized, double-blind, controlled trials involving larger sample sizes. When evaluating drugs that are intended to be used in addition to opioid therapy, the use of a placebo control is appropriate and ethical. There are two major types of randomized, double-blind, placebo-controlled trial designs in analgesic studies: parallel-group and crossover. However, if there is a significant chance that the drug will have a carryover effect or the duration of a drug’s effect is difficult to predict or is variable, a parallel-group design should be used. One should also consider that a crossover design extends the length of a trial. For the cancer pain
population progression of the disease over the course of many weeks and a consequential worsening of pain over the duration of the trial could complicate the interpretation of the results.

II.2. Short term studies

The following example will be used to illustrate a pivotal, later Phase II analgesic study: a multicentre, randomized, double-blind, placebo-controlled, parallel-group trial to evaluate the efficacy and safety of subcutaneously administered drug XXX in patients with moderate to severe cancer pain. For this example, drug XXX or placebo is administered b.i.d. (twice each day) to patients over a four day period.

II.2.A. Study Objectives

Below is an example of a set of objectives for a clinical trial to evaluate investigational drug XXX as an adjuvant to opioid therapy for patients with moderate to severe cancer pain.

Primary Objectives

a. to compare the efficacy of subcutaneous (s.c.) XXX treatment versus placebo in reducing the intensity of pain
b. to compare the safety of s.c. XXX treatment versus placebo

Secondary Objectives

a. to estimate the onset of analgesic effect of s.c. XXX
b. to estimate the time of peak analgesic effect of s.c. XXX
c. to estimate the duration of analgesic effect of s.c. XXX treatment
d. to determine whether s.c. XXX treatment reduces the need for breakthrough medication
e. to determine whether s.c. XXX treatment improves patient function.

Exploratory Objective

a. to generate preliminary information about the specificity of XXX’s analgesic action for neuropathic, visceral, and somatic pain

II.2.B. Primary endpoints

A wide variety of assessment tools are used for study endpoints to evaluate the efficacy of analgesic drugs in clinical trials and, unfortunately, there is currently no one gold standard for this very important task. Choice of pain intensity measures varies considerable across published clinical trials. These measures include variations in numeric rating scales, visual analogue scales, and verbal rating scales. All three of these scale types have been shown to be sensitive to treatment- and time-related changes in pain intensity. Pain intensity measures are also diverse with respect to whether they measure worst pain, average pain, or least pain over a specified period of time, or current pain. The Brief Pain Inventory (BPI) and other instruments use several pain intensity measures, concurrently, to estimate the degree of this dimension of pain. Measures of pain relief or patient evaluation of analgesic efficacy should also be considered in addition to pain intensity measures, since they can be effective in detecting a clinically meaningful analgesic response to treatment. However, in the literature there is considerable variation in the types of scales used to assess these two constructs. Pain intensity and pain relief measures are often concurrently used to characterize the onset, peak, and duration of an investigational drug’s analgesic effect. The post-treatment times for evaluating pain intensity and pain relief for study endpoints will vary across clinical trials and will be based on the known pharmacodynamic properties of the drug. Assessment of efficacy in the cancer pain population is often more challenging than in the non-malignant pain population. Patients with cancer often suffer with more than one pain syndrome, each syndrome presenting with one or more pain symptoms. Furthermore, cancer pain syndromes and pain symptoms can differ with respect to their underlying pathophysiological mechanisms. For trials involving the evaluation of drugs with a very specific mechanism of action, the use of one global pain
intensity measure for the primary efficacy endpoint may not be sufficiently sensitive to detect an analgesic response. A global pain intensity or pain relief measure may have limited sensitivity in situations where a particularly bothersome pain symptom responds to the drug in a clinically meaningful way, but the patient’s other pain symptoms have not responded due to differences in their underlying pathophysiological mechanisms. Unfortunately, there is currently no assessment tool specifically designed to simultaneously evaluate changes in the intensity of patients’ distinct pains to an intervention. The Neuropathic Pain Scale (NPS) comes closest to achieving this, but it is intended to be used for simultaneously evaluating multiple symptoms of neuropathic pain only (1). Thus, a patient diary can be developed for a clinical trial to include validated pain intensity and or pain relief scales that allow assessment of treatment-related changes in specific pain symptoms over time.

In further consideration of the multidimensional nature of cancer pain and to increase the sensitivity of a clinical trial in detecting clinically meaningful responses to treatment, additional measures of efficacy should be considered for study endpoints, including measures of physical functioning, emotional functioning, and opioid requirements. It is best, however, that the assessment tools chosen to measure these constructs have been validated in the cancer pain population and have been demonstrated to be sensitive to analgesic drug interventions.

See references 2 through 7 for discussions about efficacy measures for analgesic drug clinical trials. In the example below, multiple primary and secondary endpoints have been chosen to increase the sensitivity of the trial to detect a clinically meaningful response to the investigational drug and to better understand its pharmacodynamic properties. Both Day 5 and Day 8 were chosen for the endpoints in an attempt to maximize the difference in the proportion of responders in the active versus placebo groups. Based on previous studies, it is known that drug XXX produces analgesia that persists well beyond the treatment period.

The Primary endpoints include:

a. the proportion of responders in the drug XXX versus placebo groups, where a response is defined as:
   • a ≥ 30% reduction in pain intensity from baseline, on Day 5 and Day 8 post treatment, as measured by the Brief Pain Inventory – Short Form Question #3 (BPI-SF Q#3; numeric rating scale the worst pain in the last 24 h). The reduction in pain intensity must be accompanied by either a decrease or stabilization (<15% increase) in mean opioid analgesic consumption compared with baseline.

b. the proportion of clinical responders in the drug XXX versus placebo groups, where a clinical response is defined as:
   • a≥ 30% reduction in pain intensity from baseline, on Day 5 and Day 8 post treatment, as measured by the BPI-SF Q#5 (numeric rating scale for average pain in the last 24 hours) if average baseline pain intensity score is ≥4, or
   • a ≥ 30% reduction in pain intensity from baseline, on Day 5 and Day 8 post treatment, as measured by any component-specific pain scale from Patient Diary that has an average baseline pain intensity score of ≥4, or
   • a ≥ 30% reduction in pain intensity from baseline, on Day 5 and Day 8 post treatment, as measured by any component-specific pain measured by the subscales of the Neuropathic Pain Scale (NPS) that has an average baseline pain intensity score of ≥4, and
   • the patient confirms that the global pain (BPI-SF Q#5) or component-specific pain (Patient Diary or NPS) has “very much improved” or “much improved” since the start of the study (Note: these are two categories from a 7-point verbal evaluation scale asking the patient to assess how much their pain has changed since the start of the study, with response categories ranging from 1=very much improved – 7=very much worse).
Safety is assessed through the number of adverse events, the number and nature of abnormal laboratory results, and changes in 12-lead electrocardiogram, blood pressure, heart rate, respiratory rate, and SaO₂.

II.2.C. Secondary endpoints
a. the time of onset of a consistent decrease in pain intensity on the visual analogue scale (VAS) compared to baseline following the first dosing of the treatment phase
b. the post treatment day during which the greatest reduction in pain intensity (BPI-SF Q#3, BPI-SF Q#5, and/or any component-specific pain intensity rating scale) is reported by Responders, or Clinical Responders compared to baseline
c. the interval in days from the first of two consecutive days a patient reports a ≥30% reduction in pain intensity (BPI-SF Q#3, BPI-SF Q#5, and/or any component-specific pain intensity rating scale) until the reduction in pain intensity score is ≤15% compared to baseline
d. the number of treated breakthrough pain episodes post-treatment
e. the proportion of patients achieving a ≥30% of improvement in their general activity (BPI-SF Q#9A) or walking ability (BPI-SF Q#9C)

II.2.D. Exploratory endpoints
a. the proportion of patients with a neuropathic pain component or without a neuropathic pain component who are categorized as responders and non-responders

II.2.E. Study Design
Drug XXX is being evaluated by a multicentre, randomized, double-blind, placebo-controlled, parallel-group trial to evaluate the efficacy and safety of subcutaneously administered drug XXX in patients with inadequately controlled moderate to severe cancer pain. Drug XXX or placebo will be administered b.i.d. to patients over a four day period. The study duration will extend from three to 10 weeks, beginning the first day of screening until the end of patients’ analgesic response.

All concomitant medications taken by the subject from the start of the first treatment period to the completion of the Follow-Up Visit will be documented. The reported medications will be reviewed and evaluated by the Qualified Investigator or designate to determine if they affect the subject’s eligibility to continue to participate in the study.

Screening for Eligibility (Day –28 to –7)
a. Each subject will undergo screening procedures within 7-28 days prior to the baseline period. The following will be required during the screening period to determine eligibility:
b. Informed consent
c. Medical history
d. Review of concomitant medication (including analgesics)
e. Physical and neurological examinations
f. Vital signs (pulse rate, blood pressure, respiratory rate) and body weight
g. 12-lead ECG
h. Laboratory tests (haematology, clinical chemistry; urinalysis)
i. Pregnancy test for women of childbearing potential
j. Review of inclusion and exclusion criteria
k. Characterization of pain and disease:
   • primary cancer site and type
   • identification and characterization of patients’ three most bothersome pains (e.g. location, etiology, pathophysiology, quality, intensity)
l. Patient categorization in part based on the Neuropathic Pain Questionnaire (NPQ) responses (presence versus absence of a neuropathic pain component)
m. BPI-SF
Baseline Period (Day −7 to Day −1):
The duration of the baseline period will be between 5 to 7 days. The following assessments will be completed once during the baseline period, unless otherwise specified:
   a. Review of medical history from last visit to present date
   b. Review of inclusion and exclusion criteria
   c. Vital signs (pulse rate, blood pressure, respiratory rate) and body weight
   d. Review of concomitant medications
   e. Review of information related to the patients’ three most bothersome pains
   f. BPI-SF (completed daily between 18:00h-21:00h)
   g. Patient Diary (completed daily between 18:00h-21:00h):
      • intensity of three most bothersome pains
      • recording of concurrent medications, including analgesics
   h. Neuropathic Pain Scale (NPS), if applicable

Note: Calculation of average baseline pain intensity will determine final eligibility (the mean “worst” pain intensity score, calculated from the last five BPI-SF Q #3 scores recorded by the patient during the baseline period; “moderate” pain will be defined as an average score of 4-5, and “severe” defined as an average score of 6 or higher).

Treatment Phase (Days 1-4)
Patients will be admitted to the hospital (in-patient) or at the site’s care facility on a daily basis and will be allowed to leave the facility upon completion of each daily treatment session if judged appropriate by the investigator. All patients will be randomized on Day 1 to receive a s.c. injection of drug XXX or placebo twice daily for 4 consecutive days. The first dosing will be given between 8:00-10:00h, and the second dosing between 14:00-16:00h.

The following will be required prior to each dosing unless otherwise specified:
   a. Review of concomitant medication
   b. Vital signs
   c. VAS for pain intensity (VAS-PI) to help determine acute analgesic response to XXX treatment (completed prior to first dosing of Day 1 and Day 4)
   d. 12-lead ECG

Following the first dosing of Day 1 and 4, acute analgesic response will be assessed (VAS-PI) every 15 minutes for the first hour, and then every 30 minutes until the second dosing of the day. A 12-lead ECG will be completed 1-2 hours after the morning (Days 1-4) and afternoon (Days 2 and 3) dosing. The following will be required prior to discharge unless otherwise specified:
   a. Vital signs
   b. Brief neurological examination
   c. Review of adverse events/ Adverse event recording
   d. Review of around-the-clock (ATC) analgesics and breakthrough pain medications
   e. BPI-SF (completed daily between 18:00h-21:00h)
   f. Patient Diary (completed daily between 18:00h-21:00h):
      • intensity of three most bothersome pains
      • recording of concurrent medications, including analgesics

Follow-up Visits
All patients will be assessed at the clinic on follow-up Days 5, 8, and 15 and then when pain intensity returns to baseline levels. Each evening between 18:00h-21:00h during the follow-up period, patients will complete the BPI-SF, and record in their Patient Diary the intensity of the three most bothersome pains and concomitant medications. In addition, patients will record their impression of change for each
bothersome pain (from 1 = very much improved to 7 = very much worse) in their Diary on Days 5, 8 and 15. All patients will have the option to enroll in an open label extension protocol on Day 15 or later. The following will be required at each follow-up clinic visit:
   a. Physical examination
   b. Vital signs
   c. Completed BPI-SFs
   d. Completed NPSs, if applicable
   e. Review of the Patient Diary, including patient’s assessment of change
   f. Review of concomitant medication

Whenever Day 15 is the last visit, laboratory evaluations (clinical chemistry, haematology, and urinalysis), 12-lead ECG, and pregnancy test for women of childbearing potential will be completed. Patients who have experienced Severe Adverse Effects (SAEs) or Adverse Effects (AEs) that are at least possibly related to the study medication will be followed-up by telephone on Day 35 to assess their outcome.

If patient’s pain is adequately controlled in the opinion of the investigator and the subject, the following assessments will be completed by telephone or at clinic visits (required at least every two weeks) for up to 6 weeks (Day 43):
   a. Patient Diary (weekly post Day 15 until last day of study)
   b. BPI-SF (weekly post Day 15 until last day of study)
   c. Vital signs (every 2 weeks at a minimum, for a maximum of 4 weeks)
   d. NPS, if applicable (at last visit)
   e. Laboratory evaluations (clinical chemistry, haematology, and urinalysis at last visit)

II.2.E. Planned sample
Sample size calculation will be based on the primary efficacy endpoint, i.e. comparison of the proportion of responders, based on the worst pain in the last 24 hours from baseline to Day 5 between the XXX and placebo study drug groups. A total of 116 evaluable subjects will be required to detect a difference in proportion of responders between study drug groups of 20% (placebo) versus 50% (XXX) under the following assumptions: Equal numbers of subjects in the two treatment conditions; and 2-sided test, using a significance level of 0.049; and minimum power of 0.90 (90%). A 30% decrease in pain intensity is considered to be a clinically meaningful response (2). The choice of a within-patient 30% reduction in pain intensity is also based on discussion of clinical importance of changes in chronic pain intensity using an 11-point scale (3).

In a population of subjects with painful diabetic neuropathy (8), investigators observed placebo response rates ranged from 10% to 40%, with an average of 26%; it is expected that a placebo response rate for subjects with cancer pain as in the current study will be similar. It has also been shown that the effect of placebo in the treatment of refractory pain in patients with cancer was 18.1% (9). Thus, a response rate of 20% was selected as a reasonable estimate for a placebo effect for this study. If the placebo effect is larger than anticipated (e.g. 30% responder rate in the placebo group), then power is still high enough (87%) to detect the 30% difference in responder rates between the two study groups.

Assuming that 20% of the enrolled patients will discontinue the study or be withdrawn from the study, a total of 146 patients will be required for this parallel-group trial. It is planned to recruit this sample in approximately 25-30 centres across the country with a mean enrolment of 4 patients per centre. Enrolment into the screening phase of the study will be stopped when the anticipated or actual number of subjects has been achieved across all study sites.
II.2.F. Study population
Male or female subjects over 18 years of age with stable but inadequately controlled moderate to severe cancer pain of at least two weeks duration. Patients may experience visceral, somatic and/or neuropathic pain, requiring opioid administration.

II.2.G. Specific inclusion criteria
A subject will be eligible for inclusion in this study only if all of the following criteria apply:

a. Male or female 18 years of age and over;
b. In-patient or out-patient with a diagnosis of cancer;
c. Stable but inadequately controlled pain with current therapy for at least two weeks;
d. Experiencing somatic, visceral and/or neuropathic pain related to cancer;
e. Pain intensity, as assessed by BPI-SF Q#3 meets the definition of “moderate” (score of 4-5) or “severe” (score of 6-10) pain;
f. Life expectancy of > 3 months;

II.2.H. Specific exclusion criteria
For this example, the exclusion criteria have to be adapted to characteristics of the drug, e.g. the pharmacokinetics, pharmacodynamics, potential drug-drug interactions, and adverse effects. A patient will not be eligible for inclusion in this study if any of the following criteria apply:

a. Planned initiation of chemotherapy, radiotherapy, or bisphosphonates within 30 days prior to randomization;
b. Taking lidocaine, mexiletine, or other anaesthetics;
c. History of CO₂ retention, or oxygen saturation (SaO₂) <90% despite O₂ via nasal prongs;
d. Use of scopolamine, acetylcholinesterase-inhibiting drugs, beta-blockers and antiarrhythmic drugs;
e. Second or third degree heart block or prolonged QT₇ interval (corrected for rate) on screening ECG (confirmed > 470 msec on repeated occasion);
f. Known hypersensitivity to XXX and/or its derivatives;
g. Received an investigational agent within 30 days prior to screening or who is scheduled to receive an investigational drug other than XXX during the course of the study;
h. Females who are lactating or at risk of pregnancy (i.e., sexually active with fertile males and not using an adequate form of birth control);
i. Females with a positive serum pregnancy test at screening or positive urine pregnancy test on admission to study site; or
j. Any other condition that, in the opinion of the investigators, is likely to interfere with the successful collection of the measures required for the study or poses a risk to the patient.

II.2.I. Tools for assessing the endpoints
II.2.I.i. Tools to assess efficacy
The following table 1 summarizes the assessment tools chosen for this example to measure efficacy variables.

II.2.I.ii. Tools to assess safety
Safety will be evaluated through the assessment of spontaneously reported adverse events, vital signs, physical exam findings and laboratory tests. Other safety data will include 12-lead ECG measurements and brief neurological examinations.

a. Adverse events (AEs)
Adverse events will be recorded daily beginning on Day 1 up to Day 15 for non-serious AEs, and from screening up to 30 days after the last dose of the investigational product
Table 1. Tools to assess efficacy variables

<table>
<thead>
<tr>
<th>Variable</th>
<th>Assessment Tool Used to Measure Variable</th>
<th>Time of Assessment</th>
</tr>
</thead>
<tbody>
<tr>
<td>Global Pain Intensity in Last 24 Hours ('worst' &amp; ‘average’ pain) and current pain</td>
<td>BPI-SF (Q#3, Q#5, Q#6)</td>
<td>Daily through all phases of the trial until Day 15 and then weekly until the end</td>
</tr>
<tr>
<td>Pain Relief in Last 24 Hours</td>
<td>BPI-SF (Q#8)</td>
<td>Daily through all phases of the trial until Day 15 and then weekly until the end</td>
</tr>
<tr>
<td>Component-Specific Pain Intensity</td>
<td>Patient Diary (NRSs; 0=no pain, 10=pain as bad as you can imagine)</td>
<td>Daily through all phases of the trial until Day 15 and then weekly until the end</td>
</tr>
<tr>
<td>Component-Specific Pain Intensity</td>
<td>NPS (subscales)</td>
<td>Baseline (once), Days 5, 8, 15 and last day of trial</td>
</tr>
<tr>
<td>Acute Analgesic Response</td>
<td>VAS-PI</td>
<td>On Day 1 and Day 4, prior and after first dosing (every 15 minutes for the first hour, and then every 30 minutes until the second dosing)</td>
</tr>
<tr>
<td>ATC and Breakthrough Analgesic Use</td>
<td>Patient Diary</td>
<td>Continually through all stages of the trial until the end (morphine equivalents will be calculated)</td>
</tr>
<tr>
<td>Patient’s Impression of Change in Pain</td>
<td>Patient Diary (7-point categorical scale; ‘very much improved’ to ‘very much worse’)</td>
<td>Days 5, 8, 15</td>
</tr>
<tr>
<td>Time to Analgesic Response</td>
<td>BPI-SF (Q#3, Q#5) Patient Diary (component-specific pain intensity NRSs)</td>
<td>Daily through all phases of the trial until Day 15 and then weekly until the end</td>
</tr>
<tr>
<td>Duration of Analgesic Response</td>
<td>BPI-SF (Q#3, Q#5) Patient Diary (component-specific pain intensity NRSs)</td>
<td>Daily through all phases of the trial until Day 15 and then weekly until the end</td>
</tr>
<tr>
<td>Time to Peak Analgesic Response</td>
<td>BPI-SF (Q#3, Q#5) Patient Diary (component-specific pain intensity NRSs)</td>
<td>Daily through all phases of the trial until Day 15 and then weekly until the end</td>
</tr>
<tr>
<td>Walking Ability</td>
<td>BPI-SF (Q#9C)</td>
<td>Daily through all phases of the trial until Day 15 and then weekly until the end</td>
</tr>
<tr>
<td>General Activity</td>
<td>BPI-SF (Q#9A)</td>
<td>Daily through all phases of the trial until Day 15 and then weekly until the end</td>
</tr>
</tbody>
</table>

for severe AEs (SAEs). The frequency of adverse events will be tabulated and summarized according to:

- **Type**: clinical laboratory abnormalities detected in biological samples, abnormalities detected on physical and neurological examinations, adverse reactions described by the patient.
- **Severity**: mild, moderate, severe, life threatening
- **Association with treatment (causality)**: probably related, possibly related, not related, unknown.
b. Clinical laboratory evaluations
Clinical laboratory tests will be performed by a central laboratory. The following evaluations will be conducted: clinical chemistry, haematology and coagulation, urinalysis, and pregnancy test.

c. Physical examination, vital signs, body weight, height, brief neurological examination

d. 12-lead ECG
A 12-lead ECG will be recorded after 10 minutes rest in the supine position at: Screening, Pre-dose on Day 1 (randomization; triplicate), 1 hour post dose on Day 1 (triplicate), Pre-dose on Day 2 (triplicate), 1 hour post dose on Day 2 (triplicate), 1 hour post PM dose on Day 2 (triplicate), Pre-AM dose on Day 3 (triplicate), 1 hour post-PM dose on Day 3 (triplicate), Pre-AM dose on Day 4 (triplicate), 1 hour post-AM dose on Day 4 (triplicate), and Day 15.

The central reader will measure the electrocardiographic intervals manually on a computer screen using digital callipers. Each interval will be derived as a mean of three measurements taken from three consecutive QRST complexes. Mean QT and PR intervals will be used to derive the Bazett (QTcB) and Fridericia (QTcF) corrected QT intervals. A computer interpretation will be faxed to the site within 30 minutes of transmission. A cardiologist-reviewed ECG report including a full diagnostic interpretation will be faxed to the site next business day.

II.2.J. Specific criteria for early withdrawal and discontinuation
Subjects will be permitted to leave the study at any time. Subjects can be withdrawn from the study for any of the following reasons:
   a. Occurrence of a SAE
   b. Administrative reasons (e.g. sponsor decision)
   c. Withdrawal of consent
   d. Major violation of the protocol
   e. Pregnancy
   f. Non-compliance
   g. If it is of the opinion of the Qualified Investigator that it is in the best interest of the patient to discontinue

Patients discontinuing because of a SAE or because it is in the Qualified Investigator’s opinion that it is in the best interest of the patient, will be considered to have completed the study. Patients discontinuing because of withdrawn consent, a major violation of the protocol, pregnancy, or non-compliance will be non-completers and replaced if they leave prior to drug administration. If a patient is prematurely discontinued from participation in the study for any reason after drug administration, the investigator must make every effort to perform the following evaluations: physical examination, 12-lead ECG, vital signs, clinical laboratory tests including haematology, clinical chemistry, and urinalysis, adverse event assessment, and pregnancy test (females of childbearing potential). These data should be recorded in the source documentation and CRF, as they comprise an essential evaluation that should be done prior to discharging any patient from the study. These subjects will be considered to have completed the study.

In the event that a patient is prematurely discontinued from the study at any time due to an AE or SAE, the procedures stated in Section XII must be followed. The "End of Study Record" page of the CRF will be completed for any patient withdrawn from the study.

Patients who drop out of the study due to changing medical status not related to pain, clinically important changes in non-pain-related medications, or whose status meets one or more exclusion criteria will be replaced. These patients will not be considered treatment failures.
II.2.K. Data analysis

a. Analysis populations
   All efficacy analyses will be performed using the intent-to-treat principle, i.e. subjects will be analyzed based on the study drug group to which they were randomized. All safety analyses will be performed for subjects as dosed. In addition to the intent-to-treat analysis approach for efficacy (primary), if a substantial proportion of subjects (>10%) fail to complete four days of study treatment or there are a substantial number of subjects (>10%) with critical protocol violations (e.g. baseline worst pain < 4 for the BPI-SF Q#3), per-protocol analyses will be performed using subjects who complete all four days of study treatment with no critical protocol violations.

b. Significance/confidence level
   Differences will be considered statistically significant if the significance level is \( \leq 0.049 \), 2-tailed. Confidence intervals for the absolute difference between the treatment groups in the outcomes will be calculated using 95.1% confidence, 2-tailed.

c. Efficacy analysis
   The overall objective is to determine the efficacy of s.c. XXX in reducing the intensity of cancer-related inadequately controlled pain compared to placebo.

Primary efficacy analysis
   The primary efficacy analysis will be performed to compare the proportion of patients who are responders to XXX with the proportion of patients who are responders to placebo, based on the BPI-SF Q#3. Comparison of the proportion who responds in each treatment group will be made using the Mantel-Haenszel procedure, stratifying for: baseline pain pathophysiology (includes neuropathic component/does not include neuropathic component) and baseline pain intensity as determined by the average of BPI-SF Q#3 score (moderate/severe). For the primary efficacy analysis, missing data for Day 5 or Day 8 will constitute a non-responder. As a co-primary analysis, a responder analysis will be performed using the clinical responder definition of response. The same statistical analysis method will be used for the co-primary efficacy analysis as for the primary efficacy analysis. See responder and clinical responder definitions on page 4. Exploratory analyses will be performed to compare the mean change from baseline in component BPI-SF pain intensity scores and NPS scores using an analysis of covariance (ANCOVA) model, with change from baseline in pain as the dependent variable, the study group and baseline pain pathophysiology as independent variables, and mean baseline pain intensity score as a covariate. Least squares means for the study drug groups will be reported.

Secondary efficacy analysis
   (i) Determination of whether s.c. XXX treatment improves daily mobility in patients with refractory cancer pain.

   The mean change from baseline in BPI-SF Q #9A (general mobility) and BPI-SF Q #9C (walking ability) scores will be combined and compared between treatment groups using an analysis of covariance (ANCOVA) model, with change from baseline in daily mobility as the dependent variable, the study drug group and baseline pain pathophysiology stratum as independent variables, and mean baseline pain intensity score as a covariate.

   (ii) Determination of whether s.c. XXX treatment reduces the need for breakthrough medication.
   The overall number of treated breakthrough episodes per patient will be summarized. The number of treated breakthrough episodes per patient will be compared between treatment groups using a Poisson regression model with the study drug group and pain pathophysiology stratum as independent variables.

   (iii) Determination of the onset, duration, and peak of XXX analgesic effect
The onset of analgesia is defined as the first time point showing a consistent decrease in pain intensity, as measured by the VAS-PI, compared with baseline following the first XXX dosing on Day 1 and Day 4.

The duration of analgesic response is defined as the interval in days from the first of two consecutive days a patient reports a ≥30% reduction in pain intensity (BPI-SF Q#3, BPI-SF Q#5, and/or any component-specific pain intensity rating scale) until the reduction in pain intensity score is ≤15% compared to baseline or the patient confirms that the decrease in pain intensity from baseline is no longer meaningful to him or her. The duration of analgesic response may differ depending on the pain component.

Peak analgesic effect is defined as the day during which the greatest reduction in worst pain intensity occurs (BPI-SF Q#3) for responders. Time of peak analgesic effect and duration of analgesic effect will be summarized across treatment groups.

**Exploratory efficacy analysis**
The proportion of responders, using the definition of responder for the primary efficacy analysis will be summarized separately for each pain pathophysiology category (neuropathic, visceral, somatic, and mixed).

d. Safety analysis

**Adverse events**
Adverse event rates will be summarized and compared between the study drug groups for: overall incidence, incidence of related adverse events (according to the causality assessment), incidence of grade 3/4 toxicity per NCI-CTC toxicity criteria, incidence of serious adverse events (including death), and incidence of adverse events leading to discontinuation. Incidence will be counted as treatment-emergent if adverse event onset or worsening occurs after first dose of study drug. For purposes of adverse event analysis, adverse events related to pain will not be summarized.

**Vital signs assessments**
Vital sign data will be summarized at each time point and change from baseline using descriptive statistics (mean, median standard deviation, minimum, and maximum), for systolic blood pressure, diastolic blood pressure, and heart rate. Changes from baseline that exceed the limits indicated in table 2 will be tabulated for each study group.

<table>
<thead>
<tr>
<th>Table 2. Vital sign limits</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Parameters:</strong></td>
</tr>
<tr>
<td>Systolic Blood Pressure:</td>
</tr>
<tr>
<td>Diastolic Blood Pressure:</td>
</tr>
<tr>
<td>Heart Rate:</td>
</tr>
</tbody>
</table>

**ECG assessments**
ECG data will be summarized at each time point (using the mean of 3 observations observed at baseline and 3 observations observed following dosing each day) and reported as change from baseline using descriptive statistics (mean, median standard deviation, minimum, and maximum) for each day for: R-R interval, PR, QRS, QT, and QTc. Regarding QTc, the Bazett and Fredericia corrections for R-R interval will be used. Tolerance limits (Confidence Interval) for the difference between study groups for mean change in QTc at each time point will be estimated. Shift tables will be used to assess baseline versus post-baseline relation of ECG value to reference range for each study drug group, and changes in ECGs from baseline that exceed limits defined in table 3 will be tabulated for each study drug group.
Table 3. ECG reference ranges and change limits

<table>
<thead>
<tr>
<th>Parameters:</th>
<th>Change Limit:</th>
<th>Reference Range:</th>
</tr>
</thead>
<tbody>
<tr>
<td>Heart Rate:</td>
<td>Change from baseline $\geq \pm 10$ bpm</td>
<td></td>
</tr>
<tr>
<td>PR Interval:</td>
<td>Change from baseline $\geq \pm 32$ msec</td>
<td>115 msec - 200 msec</td>
</tr>
<tr>
<td>QRS Interval:</td>
<td>Change from baseline $\geq \pm 16$ msec</td>
<td>80 msec - 100 msec</td>
</tr>
<tr>
<td>QT/ QTc Interval:</td>
<td>Change from baseline $&gt; \pm 30 - 60$ msec / Change from baseline $&gt; \pm 60$ msec</td>
<td>320 msec - 470 msec</td>
</tr>
</tbody>
</table>

In addition, abnormalities in T wave/U-wave morphology will be summarized for each study drug group.

Laboratory assessments
Laboratory data will be assessed using descriptive statistics (mean, median, standard deviation, minimum, and maximum) for each of the time points for haematology, blood chemistry, and urinalysis parameters. Shift tables will be used to assess baseline versus post-baseline relation of laboratory value to reference range for each study drug group, and changes to laboratory values that are grade 3 or 4 using the NCI-CTC criteria will be tabulated for each study drug group.

II.2.L. Adverse events (AEs) and severe adverse events (SAEs)
The investigator is responsible for the detection and documentation of events meeting the definition of an AE or SAE as provided in this protocol. During the study, when there is a safety evaluation, the investigator or study site personnel will be responsible for XXX AEs and SAEs.

An adverse event (AE) is defined as an unusual and most often undesirable symptom or sign that occurs in human subjects participating in a study. Adverse Events include clinically significant abnormal laboratory values and test results, concomitant illness, accident, medical occurrence or worsening of existing medical condition that emerge during study participation.

All AEs will be recorded on the Adverse Event CRF at each assessment time or when otherwise volunteered by the subject. The investigator will actively solicit this information and assess the AEs from the subject in terms of severity and relationship to study drug. The investigator will treat the subject as medically required until the AE either resolves or becomes medically stable. This treatment may extend beyond the duration of the study. The investigator will record treatment and medications required for treatment on the appropriate CRF(s) and will provide reports of AEs to the sponsor’s clinical monitor on a regular basis during the study conduct.

The investigator will also report to the sponsor all AEs that come to his/her attention after the study termination within 30 days of the last dose of study drug(s).

III. EXAMPLES OF LANDMARK WELL DESIGNED ANALGESIC TRIALS


IV. REFERENCES

Chapter 30. Drugs Used in Osteoporosis

Kim Brixen, M.D., Ph.D.

Moustapha Kassem, M.D., Ph.D.

Bo Abrahamsen, MD, Ph.D.

Department of Endocrinology  
Odense University Hospital  
DK-5000 Odense C  
DENMARK

Corresponding Author:  
Kim Brixen, MD, PhD  
Department of Endocrinology  
Odense University Hospital  
DK-5000 Odense C  
DENMARK  
E-mail kim.brixen@ouh.fyns-amt.dk
I. INTRODUCTORY REMARKS

Osteoporosis is a disease characterized by low bone mass and deterioration of bone architecture leading to decreased bone strength and increased risk of fractures (1). This disease is implicated in the majority of low-energy fractures occurring in the elderly population. It is estimated that approximately 50% of all women and 25% of men will suffer at least one osteoporotic fracture during their lifetime (2). Such fractures often have considerable consequences for the individual due to increased morbidity and pain, loss of independence, reduced life expectancy, and reduced health-related quality of life. For the society, the burdens of hospital treatment, rehabilitation, and nursing home care, etc., in patients with fractures are enormous and expected to rise due to demographic changes. The annual costs of osteoporotic fractures and their sequels are estimated to exceed $17 billion in the U.S. alone. Thus, it is projected that the number of hip fractures will increase 4-5 times during the next 40-50 years as a consequence of the increasing population aged 65+ years. It is an additional concern that this increase will be most pronounced in the developing countries (3).

Evidently, the demand for treatment and the incentive to develop new drugs in this field are great. The use of hormonal replacement therapy (HRT) in osteoporosis was suggested in 1947 (4) but only in 1990 the first randomized, placebo-controlled trial documented the anti-fracture efficacy of a drug (etidronate) (5). Since then, several other bisphosphonates and other anti-resorptive drugs (e.g., selective estrogen receptor modifiers (SERMs)) have been developed at an ever-increasing speed. Also, anabolic agents, e.g., teriparatide and agents with apparently both anabolic and antiresorptive effects (strontium ranelate) have been introduced. Recently, large-scale randomized studies have demonstrated the anti-fracture efficacy of HRT, but also that an unfavorable ratio between effects and adverse events precludes its long-term use as prevention or therapy of osteoporosis (6), reiterating the importance of clinical testing of drugs in well-designed trials with clinically relevant outcomes. Finally, advances in molecular biology have identified an array of potential targets for new drugs such as integrins, osteoprotegerin, and osteoclast-specific chloride channels.

Both the U.S. Food and Drug Administration (FDA) (7,8) the European Medicines Evaluation Agency (EMEA) (9), and the World Health Organization (WHO) (10) have issued guidelines for the development and testing of anti-osteoporotic agents highlighting many of the specific problems that must be considered during drug development in this area. In clinical life, the number of patients with postmenopausal osteoporosis and age-related osteoporosis in women dominates the picture. Consequently, most drugs are initially tested in phase-II and –III trials in these patients while their use is only later expanded through phase-IV trials to e.g., male and glucocorticoid-induced osteoporosis.

Bone mineral density (BMD) predicts the future risk of fractures (11). Indeed, osteoporosis may be pragmatically defined as BMD below 2.5 standard deviations (SD) below the average for healthy young women (12). Treatment, however, may alter this relationship between bone mass and strength. Sodium fluoride, for example, increases BMD without decreasing the occurrence of fractures (13). Also, strontium increases BMD out of proportion with the bone accretion due to its high atomic number and higher attenuation coefficient but decreases fracture rate significantly (14). In contrast, raloxifene has little effect on BMD but still prevents fractures by yet unknown mechanisms (15). Finally, since mineralization of newly formed bone is less than that of older bone, anabolic agents such as teriparatide that speed up bone remodeling may decrease BMD initially without adversely affecting bone mass (16). Thus, preclinical studies should include evaluation of bone quality (bone mass, architecture, and strength) and clinical efficacy should be documented by fracture prevention in placebo-controlled or comparative studies.

Fractures of the hip, spine, and forearm are known to be associated with osteoporosis. However, the risk of other fractures is also increased in osteoporosis and low-energy fractures in the proximal humerus, pelvis, ribs, and ankle should rightly be considered osteoporotic (17). The incidence of vertebral fractures, however, is higher than the incidence of any of the other types of fractures and vertebral fractures are, therefore, often used as the primary outcome in clinical trials. Only about 25% of vertebral fractures identified by serial radiographs...
are associated with clinical symptoms. Back-pain may be caused by a multitude of conditions other than fracture and both painful and asymptomatic vertebral compression fractures share the same pathophysiology. Consequently, proof of efficacy may be established by reduced incidence of radiographic vertebral fractures and need not rely on clinical vertebral fractures.

Since bone structure (e.g. the relative amount of trabecular and cortical bone), pattern of loading, and hormonal responsiveness differ between skeletal sites, anti-fracture efficacy should be established both with respect to vertebral, appendicular fractures, and in particular hip fractures. If fracture prevention has been established in e.g. postmenopausal osteoporosis and if changes in BMD have been shown to reflect the reduction in fracture risk, it will generally be acceptable to rely on surrogate endpoints (i.e. BMD) in other patient groups (e.g. male or glucocorticoid-induced osteoporosis).

Throughout life, bone undergoes continuous remodeling in which bone resorption is followed by formation. To some extent, the effects of a new drug on bone remodeling may be inferred from studies of biochemical markers of bone formation (e.g. osteocalcin and alkaline phosphatase) and resorption (e.g. pyridinium cross-links and type-I collagen c-terminal telopeptide). Normally, the duration of the remodeling cycle is 4-6 months. It is clear, therefore, that clinical studies on treatment of metabolic bone diseases must be of a considerable duration to avoid misinterpretations of transient changes due to altered activation frequency or remodeling periods. Moreover, pharmacological suppression of bone resorption is usually followed some weeks or months later by decreased bone formation due to the inherent coupling between these processes in bone. A typical duration of phase-II studies is, therefore, one year and the duration of phase-III studies 3-5 years. Also, the time-frame of bone remodeling largely rules out the use of cross-over studies in this field of research (18).

An array of animal models comprising histomorphometry, measurement of bone mass (e.g. ash weight), measurement of bone strength (e.g. during compression or three-point-bending) may provide insight into most bone effects and adverse effects of a new drug such as abnormal mineralization, formation of woven bone, or interference with the healing of fractures. Bone remodeling is seen in species such as the dog but is absent or markedly less important in other species such as the rat. The FDA recommends two different animal models to be used in preclinical trials: the ovariectomized rat model (a non-remodelling species) and a remodelling species such as dog or monkey. Human studies on dynamic bone histomorphometry on bone biopsies, however, provide unique insight on alteration of bone formation, bone resorption, and bone structure during treatment. Biopsy studies are equally important as safety measures to out rule adverse effects such as woven bone, impaired mineralization (osteomalacia), and change in the mass-strength relationship. Recently, PTH-related substances have been developed for anabolic treatment of osteoporosis. In some of the animal toxicology experiments, osteosarcoma was seen in rat experiments. Additional guidelines for the development of these drugs have, therefore, been proposed by FDA (8) and authorities should be contacted for discussion of the development plan before study initiation.

A number of drugs used in osteoporosis e.g. estrogens, progestagens and SERMs may also affect other tissues than bone. During the testing of such drugs, assessment of concomitant effects on e.g. cognitive functions, cardio-vascular parameters, the mammary glands, endometrium and uterine ligaments are important.

Since osteoporosis is highly prevalent, effective drugs will see widespread use. Therefore, studies on health related quality of life are particularly important to allow estimation of pharmaco-economic implications in this field of research.

Up till now, very few head-to-head comparisons of anti-osteoporotic agents have been published and none of these have been powered to address anti-fracture-efficacy (excluding comparisons of specific drugs plus calcium and vitamin-D with calcium and vitamin-D alone). With the introduction of anabolic agents, however, both head-to-head comparisons (superiority and non-inferiority trials), combined as well as sequential treatment are to be expected.
Since osteoporosis is a disease of ageing, clinical trials should be conducted in populations similar to the target group for the approved drug. Inclusion and exclusion criteria must take into account that the patients may have multiple concomitant diseases and receive concurrent drug therapy. On the other hand, liberal criteria may lead to sizable drop-out rates due to adverse events and severely undermine conclusions (19,20). Finally, in the elderly population mortality rate is a significant factor restricting the duration of trials.

II. PHASE II STUDIES FOR REGISTRATION OF NEW DRUGS

II.1. Outline of a typical development plan

Phase-II studies should include placebo-controlled, double-blind, parallel group studies of a duration of 12 months (7) in patients with post-menopausal osteoporosis or osteopenia, e.g. long-term studies. It has been suggested that PTH-related compounds should be confined to patients with manifest osteoporosis (i.e. T-score at the hip or lumbar spine of less than -2.5 and at least one osteoporotic fracture) (8). Several dose-levels should be explored to establish the minimal effective dosage and a dose-response curve. Phase-II studies typically involve 100-400 patients and 1-10 centers.

II.2. Long-term studies

II.2.A. Objectives
The aim of phase-II studies is to document an increase or prevention of a decline in BMD in patients with osteoporosis or at risk for osteoporosis, respectively. Also, the study should demonstrate the safety of the new drug.

II.2.B. Primary endpoints
Primary endpoints are BMD of the lumbar spine and hip.

II.2.C. Secondary endpoints
Secondary endpoints include BMD of the distal forearm and whole body. Moreover, biological actions on e.g. serum calcium, PTH, 25-OH-vitamin-D, 1,25-(OH)2-vitamin-D, biochemical markers of bone formation and resorption, calcium absorption and bone histomorphometry are usually investigated in all or some of the participants in phase-II studies. Data on fractures should be collected as a safety parameter.

II.2.D. Study design
The study design should be randomized, double-blind and placebo-controlled with several parallel groups treated with different dosages the study drug, although, the use of placebo in osteoporosis has recently been challenged (see below). As discussed above, cross-over studies are usually inappropriate. Block randomization should be used to ensure similar numbers in the groups. Measurements will typically be performed at baseline, at 3, 6, and 12 months.

Concomitant therapy
a. Adequate intake of vitamin-D (400-800 IU/day) and calcium (i.e. 1500 mg/day) by the diet or supplementation should be assured. Administration of supplementation is often advisable. If dietary intake is preferred, reassessment during the study may be necessary.

b. Estrogens and progestagens should not be used unless they are part of the treatment protocol itself.

II.2.E. Planned sample
The planned sample size should be calculated from the expected increase in BMD (see below).

II.2.F. Study population
Phase-II studies usually comprise post-menopausal women with osteoporosis or osteopenia.
II.2.G. Specific inclusion criteria
a. Post-menopausal women with osteopenia (i.e. low BMD with a T-score from -1 to -2.5)
b. Post-menopausal women with osteoporosis (i.e. low BMD with T-score less than -2.5)
c. Age above 50 years (in studies on osteopenia) or 55 years (in studies on osteoporosis).

II.2.H. Specific exclusion criteria
a. Vertebral fractures of vertebral bodies L1-L4 that may preclude measurement of BMD in the spine may be considered reason for exclusion.
b. Dementia
c. Metabolic bone diseases other than osteoporosis (e.g. hypocalcemia, hypercalcaemia, ostegenesis imperfecta, malabsorption, hyperparathyroidism, renal osteodystrophy, osteomalacia or Paget’s disease)
d. Breast cancer or other malignant disease. Usually basal and squamous cell carcinoma cured by surgery or irradiation are allowed.
e. Past or present venous thromboembolism or cerebral vascular accidents (in studies on estrogens, progestagens and SERMs)
f. Liver and kidney disease
g. Alcohol or drug abuse
h. Treatment at any time with fluoride or bisphosphonates
i. Treatment within 6-12 months with calcitonin, HRT, SERMs, systemic glucocorticoids, lithium or anti-convulsants.
j. Medication that may affect bone metabolism (e.g. phosphate binding antacids, many diuretics, adrenal or anabolic steroids, heparin, anticonvulsants, supplements of vitamin D or A in excess of RDAs).
k. Patients with Paget’s disease of bone (or unexplained elevated serum levels of alkaline phosphatase) should not participate in trials on PTH-related drugs due to the inherent risk of osteosarcoma in this disease.

II.2.I. Tools for assessing endpoints
Primary endpoints
BMD is usually assessed by dual energy X-ray absorptiometry (DEXA) at the spine (AP-projection), total hip, femoral neck, distal forearm and whole body. Although DEXA scanners are available from many manufacturers, most multi-centre studies are performed using equipment from the same or a few manufacturers. Long-term stability must be ensured by daily measurements of phantoms. Cross-calibration between centers must be ensured using a portable phantom measured at each site at regular intervals. Finally, the analysis of DEXA-scans should be performed according to the same protocol or ideally at a central facility.

In some cases, such as drugs with discrepant effects on cortical and trabecular bone, BMD should also be assessed by quantitative CT in order to distinguish between effects on trabecular and cortical bone.

Secondary endpoints
Serum and urinary levels of calcium, phosphate and alkaline phosphatase may be assessed by standard methods. Serum levels of PTH, 25-OH-vitamin-D, 1,25-(OH)₂-vitamin-D and biochemical markers of bone turnover may be determined in the same laboratory. Analysis of bone biopsies requires a specialized laboratory skilled in histomorphometry.

II.2.J. Specific criteria for early withdrawal and discontinuation
In studies on patients with osteoporosis, withdrawal of patients suffering osteoporotic fractures during the study should be considered.

II.2.K. Data analysis method
BMD should be analysed following transformation to “rate of change” or “change from baseline”. Sample size should depend on the specific aim and population. The power of the study on anti-resorptive drugs should be sufficient to demonstrate a difference between the placebo and study drug groups in change in BMD of e.g. 2-3% at one year with an alpha of 0.05 and beta of 0.20 or less.
II.2.L. Special considerations
Regarding trials on PTH-related drugs, it has been proposed by the FDA that the occurrence of osteosarcoma in the rat toxicology studies must be mentioned in the patient information (8).

III. PHASE III STUDIES FOR REGISTRATION OF NEW DRUGS FOR POST-MENOPAUSAL OSTEOPOROSIS

III.1. Outline of a typical development plan
During phase III development at least one large multi-centre double-blind, placebo-controlled study is performed in patients with post-menopausal osteoporosis (“treatment”) or osteopenia (“prevention”), usually with one to three dosages of the study medication.

The duration is required to be at least 3 years for therapeutic and 2 years for prevention studies, but may be up to 5 years (long-term studies). The primary outcome must be either vertebral or appendicular fractures. Drugs classes documented not to change the relationship between BMD and fracture risk such as estrogen may be approved on the basis of studies with BMD as an endpoint in the US (7). If anti-fracture efficacy has been demonstrated in the “treatment” arm, BMD may be a sufficient endpoint in the “prevention” arm.

Comparative studies may be performed. In such case, however, US authorities demand that the study drug should be superior to standard therapy (7). Phase-III studies typically involve 1,000-10,000 patients at 10 to 300 centers.

Selection of patients with high risk of future fractures on the basis of age, a recent osteoporotic fracture and low BMD will ensure a high rate of events allowing demonstration of efficacy.

It may be argued that placebo-controlled trials are unethical since efficacious treatment is available. However, since no treatment is fully effective, the development of new drugs is highly warranted and placebo-controlled trials are the most efficient means to test a new drug, it may be argued that placebo-controlled trials should be continued. The participants, however, must be fully informed on all possible treatments. A detailed discussion of this dilemma on the basis of two separate conferences devoted to his issue has recently been the reported (21).

III.2. Long-term studies

III.2.A. Objectives
The aim of the phase III trials is to prove anti-fracture efficacy and safety of the study drug.

III.2.B. Primary endpoints
The primary outcome is either incident vertebral compression fractures, appendicular fractures or both.

III.2.C. Secondary endpoints
If the primary endpoint is vertebral fractures, appendicular fractures may be assessed as secondary endpoint. In virtually all studies, BMD at the spine, hip, and distal forearm are secondary endpoints. In contrast, ultrasound properties alone are not sufficient. Body height usually decreases with increasing numbers of vertebral fractures and should be monitored. Bone turnover as assessed by biochemical markers and calcium metabolism (serum levels of PTH, vitamin-D, calcium ion, and phosphate).

Back pain (days of disability or bed-rest) may be assessed and is important for cost-efficacy studies.
III.2.D. Safety endpoints
In addition to usual safety parameters, histomorphometric assessment of bone mineralization should be considered in sub-sample of the patients.

Moreover, some drugs e.g. estrogens, progestagens and SERMs may have a number of non-bone effects. In such drugs, mammograms, gynecological data (PAP-smears, UL of the uterine cavity, endometrial biopsy), cognitive tests, parameters of lipid metabolism and gastroscopy may be relevant safety parameters.

III.2.E. Study design
These studies should be performed with randomized, placebo-controlled, parallel groups. Cross-over studies are inappropriate. Block randomization should be used to ensure similar numbers in the groups.

Concomitant therapy:
   a. Adequate intake of vitamin-D (i.e. 400-800 IU/day) and calcium (i.e. 1500 mg/day) by diet or supplementation should be assured. Administration of supplementation is often advisable. If dietary intake is preferred, reassessment during the study may be necessary.
   b. Estrogens and progestagens should not be used unless they are part of the treatment protocol.

III.2.F. Planned sample
The sample size should be calculated on the basis on the expected fracture rate in the placebo group and the expected relative risk reduction with active treatment.

III.2.G. Study population
Post-menopausal women with manifest osteoporosis (i.e. with prevalent fractures), osteoporosis (i.e. BMD<-2.5), or osteopenia (BMD from -1 to -2.5) usually comprise the study population.

III.2.H. Specific inclusion criteria
Depending on whether prevention or therapy is the aim of the study, inclusion criteria generally comprise one of the following:
   a. BMD of the lumbar spine and/or hip from 1.0 to 2.5 SD below normal mean for young healthy women (osteopenia).
   b. BMD of the lumbar spine and/or hip 2.5 SD below normal mean for young healthy women (osteoporosis).
   c. One or more vertebral fractures defined as at least 20% reduced vertebral height (manifest osteoporosis).
   d. At least one recent low-energy fracture in the spine or appendicular skeleton.

In addition, patients should fulfill the following criteria:
   a. Post-menopausal women, e.g. 5 years (therapeutic studies) or 1-3 years (prevention studies) after the last menstruation or ovariectomized women with post-menopausal levels of serum FSH and estradiol. In studies on e.g. SERMs it may be required that the participants are more than 2 years after the menopause.
   b. Age from 50 (55) years to 80 (85) years.
   c. Informed consent

III.2.I. Specific exclusion criteria
   a. Vertebral fractures of more than two of the vertebral bodies L1-L4, that may preclude measurement of BMD in the spine, may be criteria for exclusion.
   b. Dementia
c. Metabolic bone diseases other than osteoporosis (e.g. hypocalcemia, hypercalcaemia, osteogenesis imperfecta, malabsorption, hyperparathyroidism, renal osteodystrophy, osteomalacia, or Paget’s disease)
d. Breast cancer or other malignant disease. Usually patients with basal and squamous cell carcinoma cured by surgery or irradiation are usually allowed to participate.
e. Past or present venous thromboembolism or cerebral vascular accidents (in studies on SERM)
f. Liver, kidney disease
g. Vitamin-D deficiency
h. Alcohol or drug abuse
i. Treatment at any time with fluoride or bisphosphonates
j. Treatment within 6-12 months with calcitonin, HRT, SERMs, systemic glucocorticoids, lithium or anti-convulsants.
k. Medication that may affect bone metabolism (e.g. phosphate binding antacids, many diuretics, adrenal or anabolic steroids, heparin, anticonvulsants, supplements of vitamin D or A in excess of RDAs.
l. Estrogens and progestins should not be used unless they are part of the treatment protocol.

If phase-III studies are considered in men, hypogonadism should be diagnosed and treated before study entry. Similarly, a run-in period with treatment of mild vitamin-D deficiency may be part of studies in elderly patients.

**III.2.J. Tools for assessing endpoints**

*Primary endpoints*
Serial X-rays (lateral projection) of the thoracic and lumbar spine, typically every year. Images are usually read centrally using semi-quantitative scale or morphometry (measurement of the height of the individual vertebral bodies) by one or a few observers blinded not only to the treatment allocation but also the sequence of images. Fractures are typically defined as 20% or more reduced posterior, middle, or anterior vertebral height.

*Secondary endpoints*
Dual energy X-ray absorptiometry are typically measured at the spine (AP-projection), total hip, femoral neck and distal forearm. Although, DEXA machines are available from many manufacturers, most multi-centre studies are performed using Lunar, Hologic, or Norland machines only and stability ensured with measurements of phantoms, cross-calibration ensured using a portable phantom. Appendicular fractures should be documented by X-ray reports in all cases.

**III.2.K. Specific criteria for early withdrawal and discontinuation**
Lack of efficacy as indicated by significant bone loss at the spine or hip or the occurrence of one or more new vertebral, hip or forearm fractures should allow withdrawal or additional active treatment.

Similarly, specific adverse events such as venous thrombo-embolism, ovarian, uterine or breast cancer may constitute reason for withdrawal depending on the drug in question.

**III.2.L. Data analysis method**
All statistical analysis including sub-group analysis should be pre-planned. Anti-fracture efficacy should be addressed in predefined subgroups such as a) women with prevalent fractures and b) women with low BMD but no prevalent fractures at baseline. Data should be subjected to both intention-to-treat and per-protocol analysis. Data on each type of fracture (e.g. vertebral, radius and femoral) should reported separately and together. New and worsening fractures should be reported separately in terms of “number of patients” with the event. Moreover, life-table analyses should be employed for “time to first new fracture”. The use of fracture rate may produce a skewed estimate and is not acceptable since fracture events are not independent; a patients sustaining one fracture is more likely to suffer a second fracture.
BMD should be analyzed following transformation to “rate of change” or “change from baseline”. Interim analyses are generally not encouraged and should be discussed with authorities if planned.

The power of the study should be sufficient to demonstrate a difference between the placebo and study drug groups in fracture incidence of 20-30% with an alpha of 0.05 and beta of 0.20 or less.

IV. PHASE IV STUDY TO EVALUATE THE EFFECTIVENESS AND SAFETY OF A NEW DRUG

IV.1. Outline of a typical development plan

Approximately 20% of the patients with osteoporosis develop this condition secondary to other disease or treatment. Efficacy of a drug under development in male, glucocorticoid-induced, immobilisation, inflammatory bowel disease and transplantation-induced osteoporosis is usually only assessed in phase-IV studies once anti-fracture efficacy has been established in phase-III trials and may rely on surrogate endpoints, i.e. BMD. Also, phase-IV trials may also address new modes of administration or dosing schedules (e.g., the use of suppositories or transdermal formulations, or administration once-weekly or once-a-month), combined (e.g., combined bisphosphonates and HRT, combined PTH and bisphosphonates) or sequential use of previously registered drugs (e.g., bisphosphonates followed by PTH). Finally, phase-IV studies are used to evaluate the duration of effect after termination of treatment.

IV.2. Long-term study

IV.2.A. Objectives
The objectives of phase-IV studies are usually to demonstrate efficacy similar to that of established in phase-III trials using surrogate endpoints.

IV.2.B. Primary endpoints
In most studies, BMD of the spine and hip are primary endpoints. In other studies the endpoint is patient’s preference (e.g., in studies on different modes of administration).

IV.2.C. Secondary endpoints
Incidence of osteoporotic fractures is often reported as “adverse events” in phase-IV studies.

IV.2.D. Exploratory endpoints
Drug targeting algorithms, i.e. identification of subgroups of patients likely to benefit by treatment with the study drug, may be explored in phase-IV studies.

IV.2.E. Study design
In most cases, studies will be double-blind and placebo-controlled. A design with parallel groups may be employed.

Concomitant therapy:
Considerations regarding concomitant therapy are similar to those regarding phase-III trials.

IV.2.F. Planned sample
Considerations regarding sample size are similar to those mentioned regarding phase-III studies.

IV.2.G. Study population
The study population should mirror the particular indication under investigation.

IV.2.H. Specific inclusion criteria
Considerations regarding inclusion criteria are similar to those regarding phase-III trials.
IV.2.I. Specific exclusion criteria
Considerations regarding exclusion criteria are similar to those regarding phase-III trials.

IV.2.J. Tools for assessing endpoints

Primary endpoints
The tools for assessing primary endpoints are similar to those mentioned above in phase-III trials.

Secondary endpoints
The tools for assessing secondary endpoints are similar to those mentioned above in phase-III trials.

IV.2.K. Measure of endpoints
The measurement of endpoints is similar to those mentioned above in phase-III trials.

IV.2.L. Specific criteria for early withdrawal and discontinuation
Considerations regarding early withdrawal are similar to those regarding phase-III trials.

IV.2.M. Data analysis method
Considerations regarding data analysis methods are similar to those regarding phase-III trials.

V. EXAMPLES OF LANDMARK WELL DESIGNED PHASE II TRIALS


VI. EXAMPLES OF LANDMARK WELL DESIGNED PHASE III TRIALS


VII. EXAMPLES OF LANDMARK WELL DESIGNED TRIALS


VIII. SUGGESTED READINGS


