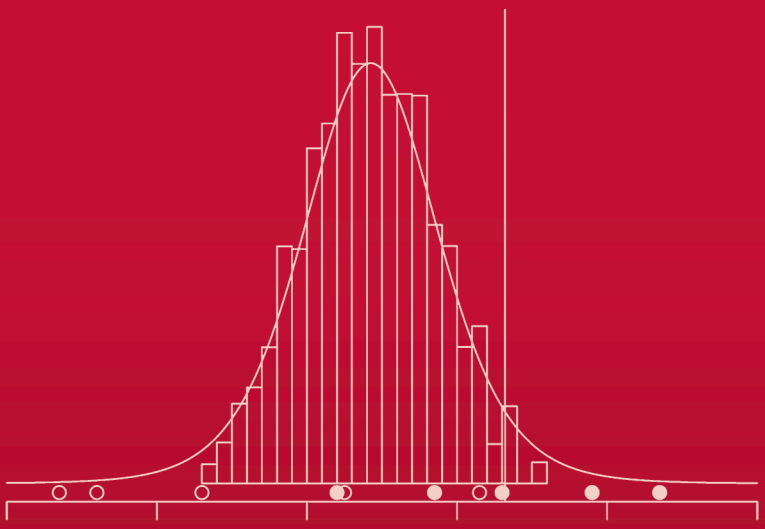
Texts in Statistical Science

Introduction to Statistical Methods for Clinical Trials



Edited by
Thomas D. Cook
David L. DeMets

Chapman & Hall/CRC
Taylor & Francis Group

Introduction to Statistical Methods for Clinical Trials

CHAPMAN & HALL/CRC
Texts in Statistical Science Series

Analysis of Failure and Survival Data
P. J. Smith

The Analysis and Interpretation of Multivariate Data for Social Scientists
D.J. Bartholomew, F. Steele, I. Moustaki, and J. Galbraith

The Analysis of Time Series — An Introduction, Sixth Edition
C. Chatfield

Applied Bayesian Forecasting and Time Series Analysis
A. Pole, M. West and J. Harrison

Applied Nonparametric Statistical Methods, Fourth Edition
P. Sprent and N.C. Smeeton

Applied Statistics — Handbook of GENSTAT Analysis
E.J. Snell and H. Simpson

Applied Statistics — Principles and Examples
D.R. Cox and E.J. Snell

Bayes and Empirical Bayes Methods for Data Analysis, Second Edition
B.P. Carlin and T.A. Louis

Bayesian Data Analysis, Second Edition
A. Gelman, J.B. Carlin, H.S. Stern and D.B. Rubin

Beyond ANOVA — Basics of Applied Statistics
R.G. Miller, Jr.

Computer-Aided Multivariate Analysis, Fourth Edition
A.A. Afifi and V.A. Clark

A Course in Categorical Data Analysis
T. Leonard

A Course in Large Sample Theory
T.S. Ferguson

Data Driven Statistical Methods
P. Sprent

Decision Analysis — A Bayesian Approach
J.Q. Smith

Elementary Applications of Probability Theory, Second Edition
H.C. Tuckwell

Elements of Simulation
B.J.T. Morgan

Epidemiology — Study Design and Data Analysis, Second Edition
M. Woodward

Essential Statistics, Fourth Edition
D.A.G. Rees

Extending the Linear Model with R: Generalized Linear, Mixed Effects and Nonparametric Regression Models
J.J. Faraway

A First Course in Linear Model Theory
N. Ravishanker and D.K. Dey

Generalized Additive Models: An Introduction with R
S. Wood

Interpreting Data — A First Course in Statistics
A.J.B. Anderson

An Introduction to Generalized Linear Models, Second Edition
A.J. Dobson

Introduction to Multivariate Analysis
C. Chatfield and A.J. Collins

Introduction to Optimization Methods and Their Applications in Statistics
B.S. Everitt

Introduction to Randomized Controlled Clinical Trials, Second Edition
J.N.S. Matthews

Introduction to Statistical Methods for Clinical Trials
Thomas D. Cook and David L. DeMets

Large Sample Methods in Statistics
P.K. Sen and J. da Motta Singer

Linear Models with R
J.J. Faraway

Markov Chain Monte Carlo — Stochastic Simulation for Bayesian Inference, Second Edition
D. Gamerman and H.F. Lopes

Mathematical Statistics
K. Knight

Modeling and Analysis of Stochastic Systems
V. Kulkarni

Modelling Binary Data, Second Edition
D. Collett

Modelling Survival Data in Medical Research, Second Edition
D. Collett

Multivariate Analysis of Variance and Repeated Measures — A Practical Approach for Behavioural Scientists
D.J. Hand and C.C. Taylor

Multivariate Statistics — A Practical Approach
B. Flury and H. Riedwyl

Practical Data Analysis for Designed Experiments
B.S. Yandell

Practical Longitudinal Data Analysis
D.J. Hand and M. Crowder

Practical Statistics for Medical Research
D.G. Altman

Probability — Methods and Measurement
A. O'Hagan

Problem Solving — A Statistician's Guide, Second Edition
C. Chatfield

Randomization, Bootstrap and Monte Carlo Methods in Biology, Third Edition
B.F.J. Manly

Readings in Decision Analysis
S. French

Sampling Methodologies with Applications
P.S.R.S. Rao

Statistical Analysis of Reliability Data
M.J. Crowder, A.C. Kimber, T.J. Sweeting, and R.L. Smith

Statistical Methods for Spatial Data Analysis
O. Schabenberger and C.A. Gotway

Statistical Methods for SPC and TQM
D. Bissell

Statistical Methods in Agriculture and Experimental Biology, Second Edition
R. Mead, R.N. Curnow, and A.M. Hasted

Statistical Process Control — Theory and Practice, Third Edition
G.B. Wetherill and D.W. Brown

Statistical Theory, Fourth Edition
B.W. Lindgren

Statistics for Accountants
S. Letchford

Statistics for Epidemiology
N.P. Jewell

Statistics for Technology — A Course in Applied Statistics, Third Edition
C. Chatfield

Statistics in Engineering — A Practical Approach
A.V. Metcalfe

Statistics in Research and Development, Second Edition
R. Caulcutt

Survival Analysis Using S—Analysis of Time-to-Event Data
M. Tableman and J.S. Kim

Time Series Analysis
Henrik Madsen

The Theory of Linear Models
B. Jørgensen

Texts in Statistical Science

Introduction to Statistical Methods for Clinical Trials

Edited by
Thomas D. Cook
David L. DeMets

Chapman & Hall/CRC
Taylor & Francis Group
Boca Raton London New York

Chapman & Hall/CRC is an imprint of the
Taylor & Francis Group, an informa business

CRC Press
Taylor & Francis Group
6000 Broken Sound Parkway NW, Suite 300
Boca Raton, FL 33487-2742

Version Date: 20110713

International Standard Book Number-13: 978-1-4200-0996-5 (eBook - PDF)

Contents

List of figures

List of tables

Preface

This text is intended to be an introduction to statistical methods for clinical trials targeted to first- or second-year graduate students in statistics or biostatistics. It arose out of a very successful course taught at the University of Wisconsin in the joint statistics/biostatistics training program for the past 25 years. The structure is similar to a text by Friedman, Furberg, and DeMets entitled *Fundamentals of Clinical Trials* but with technical material included.

The topics are based on our collective experience in the design, conduct, and analysis of clinical trials in a variety of disease areas. The material is presented from a frequentist statistical perspective although some of the topics could also have a Bayesian presentation.

The chapters have been contributed by members of the Department of Biostatistics and Medical Informatics at the University of Wisconsin School of Medicine and Public Health. The editors, who are also chapter authors, have given organization to the text and provided extensive review and input to all the chapters as they have evolved. The authors of individual chapters have interest, experience, and expertise in the topics discussed and are identified on a separate authors page. The editors endorse and take full responsibility for all the material appearing in this book.

There is no ideal sequence to the topics we have selected but we have tried to follow the thought process used in the development of a protocol. Consequently, many of the chapters are interrelated and it may be necessary for the reader to occasionally "read ahead." For example, in order to understand some discussion in the sample size chapter, the reader may have to skip forward to chapters on survival analysis or repeated measures. Throughout the text, the authors have cross-referenced other chapters where further detail on a given topic may be found.

We have also tried throughout to remain consistent with three overarching philosophical approaches described in Chapter 2, *Defining the Question*. The first is our strong belief that the design, conduct, and analysis of randomized control trials (RCTs) should adhere, to the extent possible, to the *intent-to-treat* (ITT) principle. To our dismay, the use of "per-protocol" or "on-treatment" analyses is far too widespread. Recent events illustrate how departures from this principle can lead to confusing and misleading information. While valid alternatives to ITT may be available in a few very simple situations, it is our belief that, in the vast majority of cases, there are currently no valid, practical alternatives.

Our second overarching philosophical viewpoint is that RCTs are primarily

hypothesis testing instruments. While inference beyond simple tests of the primary and secondary hypotheses is clearly essential for a complete understanding of the results, we note that virtually all design features of an RCT are formulated with hypothesis testing in mind. Some of the material, especially in Chapter 8, *Longitudinal Data*, and Chapter 9, *Quality of Life*, is unavoidably focused on complex model-based inference. Even in the simplest situations, however, estimation of a "treatment effect" is inherently model-based, dependent on implicit model assumptions, and the most well conducted trials are subject to biases that require that point estimates and confidence intervals be viewed cautiously. Inference beyond the population enrolled and treated under the circumstances of a carefully conducted trial is precarious—while it may be safe to infer that treatment A is superior to treatment B based on the result of RCTs (a conclusion based on a hypothesis test), it is less so to infer that the *size* of the effect seen in an RCT (even if could be known without error) would be realized once a treatment is adopted in common practice.

Thus, the third overarching philosophical perspective that we adopt is that the results of RCTs are best understood through the application of sound statistical principles, such as ITT, followed by interpretation rooted in clinical and scientific understanding. By this we mean that, while many scientific questions emerge in the analysis of trial data, a large proportion of these have no direct statistical answer. Nonetheless, countless "exploratory" analyses are performed, many of which deviate from sound statistical principles and either do not contribute to scientific understanding, or are in fact misleading. Our belief is that researchers, and especially statisticians, need to understand the inherent limitations of clinical studies and thoughtfully conduct analyses that best answer those questions for which RCTs are suited.

Chapter 1 introduces the clinical trial as a research method and many of the key issues that must be understood before the statistical methods can take on meaning. While this chapter contains very little technical material, many of the issues have implications for the trial statistician and are critical for statistical students to understand. Chapter 12, the last chapter, discusses the importance of the manner in which results of a trial are presented. In between, there are 10 chapters presenting various statistical topics relevant to the design, monitoring, and analysis of a clinical trial.

The material presented here is intended as an introductory course that should be accessible to masters degree students and of value to PhD graduate students. There is more material than might be covered in a one-semester course and so careful consideration regarding the amount of detail presented will likely be required.

The editors are grateful to our department colleagues for their contributions, and to a graduate student, Charlie Casper, who served as editorial assistant throughout the development of the text. His involvement was instrumental in its completion. In addition to the editors and contributors, we are grateful for helpful comments that have been received from Adin-Cristian Andrei, Murray Clayton, Mary Foulkes, Anastasia Ivanova, and Scott Diegel.

We also note that most of the data analysis and the generation of graphics in this book was conducted using R (R Development Core Team 2005) statistical software.

Thomas Cook
David DeMets

Madison, Wisconsin
July, 2007

Author Attribution

In preparing this text, the faculty listed below in the Department of Biostatistics and Medical Informatics took responsibility for early drafts of each chapter, based on their expertise and interest in statistical methodology and clinical trials. The editors, in addition to contributing to individual chapters, revised chapters as necessary to provide consistency across chapters and to mold the individual chapters into a uniform text. Without the contribution of these faculty to drafting these chapters, this text would not have been completed in a timely fashion, if at all.

Chapter				
1	Introduction to Clinical Trials	DeMets	Fisher	
2	Defining the Question	Cook	Casper	
3	Study Design	DeMets	Chappell	Casper
4	Sample Size	Cook		
5	Randomization	Casper	Chappell	
6	Data Collection and Quality Control	Bechhofer	Feyzi	Cook
7	Survival Analysis	Cook	Kim	
8	Longitudinal Data	Lindstrom	Cook	
9	Quality of Life	Eickhoff	Koscik	
10	Data Monitoring and Interim Analysis	Kim	Cook	DeMets
11	Selected Issues in the Analysis	DeMets	Cook	Roecker
12	Closeout and Reporting	DeMets	Casper	

Contributors

Robin Bechhofer, BA Researcher
T. Charles Casper, MS Research Assistant & Graduate Student
Richard Chappell, PhD Professor of Biostatistics and Statistics
Thomas Cook, PhD Senior Statistical Scientist
David DeMets, PhD Professor of Biostatistics and Statistics
Jens Eickhoff, PhD Statistical Scientist
Jan Feyzi, MS Researcher
Marian R. Fisher, PhD Research Professor
Kyungmann Kim, PhD Professor of Biostatistics
Rebecca Koscik, PhD Statistical Scientist
Mary Lindstrom, PhD Professor of Biostatistics
Ellen Roecker, PhD Senior Statistical Scientist

CHAPTER 1

Introduction to Clinical Trials

Clinical trials have become an essential research tool for the evaluation of the benefit and risk of new interventions for the treatment or prevention of disease. Clinical trials represent the experimental approach to clinical research. Take, for example, the modification of risk factors for cardiovascular disease. Large observational studies such as the Framingham Heart Study (Dawber et al. 1951) indicated a correlation between high cholesterol, high blood pressure, smoking, and diabetes with the incidence of cardiovascular disease. Focusing on high cholesterol, basic researchers sought interventions that would lower serum cholesterol. While interventions were discovered that lowered cholesterol, they did not demonstrate a significant reduction in cardiovascular mortality.[1] Finally, in 1994, a trial evaluating a member of the statin class of drugs demonstrated a reduction in mortality (Scandanavian Simvistatin Survival Study 1994). With data from well-controlled clinical trials, an effective and safe intervention was identified. Sometimes interventions can be adopted without good evidence and even become widely used. One case was the use of hormone replacement therapy (HRT) that is used to treat symptoms in postmenopausal women and is also known to reduce bone loss in these women, leading to reduced bone fracture rates. HRT also reduces serum cholesterol leading to the belief that it should also reduce cardiovascular mortality and morbidity. In addition, large observational studies have shown lower cardiovascular mortality for women using HRT than for those not using HRT (Barrett-Connor and Grady 1998). These observations led to a widespread use of HRT for the prevention of cardiovascular mortality and morbidity as well as the other indications. Subsequently, two trials evaluated the benefits of HRT in postmenopausal women: one trial in women with existing cardiovascular disease and a second without any evident disease. The first trial, known as HERS, demonstrated no benefit and suggested a possible risk of thrombosis (i.e., blood clots) (Grady et al. 1998). The second trial, known as the Women's Health Initiative, or WHI, demonstrated a harmful effect due to blood clotting and no cardiovascular benefit.[2] These trials contradicted evidence derived from non-randomized trials and led to a rapid decline in the use of HRT for purposes of reducing cardiovascular disease. HRT is still used when indicated for short-term symptom relief in postmenopausal women.

[1] The Coronary Drug Project Research Group (1975), The Lipid Research Clinics Program (1979)

[2] Writing Group for the Women's Health Initiative Randomized Controlled Trial (2002)

Incomplete understanding of the biological mechanism of action can sometimes limit the adoption of potentially effective drugs. A class of drugs known as *beta-blockers* was known to be effective for lowering blood pressure and reducing mortality in patients suffering a heart attack. Since these drugs lower blood pressure and lower heart rate, scientists believed these drugs should not be used in patients with heart failure. In these patients, the heart does not pump blood efficiently and it was believed that lowering the heart rate and blood pressure would make the problem worse. Nonetheless, a series of trials demonstrated convincingly an approximate 30% reduction in mortality.[3] An effective therapy was ignored for a decade or more because of belief in a mechanistic theory without clinical evidence. Thus, clinical trials play the critical role of sorting out effective and safe interventions from those that are not.

The fundamental principles of clinical trials are heavily based on statistical principles related to experimental design, quality control, and sound analysis. No analytical methods can rescue a trial with poor experimental design and the conclusions from a trial with proper design can be invalid if sound analytical principles are not adhered to. Of course, collection of appropriate and high quality data is essential. With this heavy reliance on statistical principles, a statistician must be involved in the design, conduct, and the final analyses phases of a trial. A statistician cannot wait until after the data have been collected to get involved with a clinical trial. The principles presented in this text are an introduction to important statistical concepts in design, conduct, and analysis.

In this chapter, we shall briefly describe the background and rationale for clinical trials, and their relationship to other clinical research designs as well as defining the questions that clinical trials can best address. For the purposes of this text, we shall define a clinical trial to be a prospective trial evaluating the effect of an intervention in humans. The intervention may be a drug, biologic (blood, vaccine, and tissue, or other products, derived from living sources such as humans, animals, and microorganisms), device, procedure, or genetic manipulation. The trial may evaluate screening, diagnostic, prevention, or therapeutic interventions. Many trials, especially those that attempt to establish the role of the intervention in the context of current medical practice, may have a control group. These and other concepts will be further discussed in this chapter and in more detail in subsequent chapters.

First, the historical evolution of the modern clinical trial is presented followed by a discussion of the ethical issues surrounding the conduct of clinical research. A brief review of various types of clinical research is presented emphasizing the unique role that clinical trials play. The rationale, need, and the timing of clinical trials are discussed. The organizational structure of a clinical trial is key to its success regardless of whether the trial is a single-center trial or a multicenter trial. All of the key design and conduct issues must be described in a research plan called a *trial protocol*.

[3] The International Steering Committee on Behalf of the MERIT-HF Study Group (1997), Krum et al. (2006), Packer et al. (2001)

1.1 History and Background

The era of the modern day clinical trial began in the post–World War II period, beginning with two trials in the United Kingdom sponsored by the Medical Research Council (1944). The first of these trials was conducted in 1944 and studied treatments for the common cold. The second trial, conducted in 1948, evaluated treatments for tuberculosis, comparing streptomycin to placebo. Hill (1971) incorporated many features of modern clinical trials such as randomization and a placebo-treated control group into this trial (Medical Research Council 1948).

In the United States, the era of the modern clinical trial probably began with the initiation of the Coronary Drug Project (CDP)[4] in 1965. The CDP was sponsored by the National Heart Institute (later expanded to be the National Heart, Lung, and Blood Institute or NHLBI), one of the major institutes in the National Institutes of Health (NIH). This trial compared five different lipid-lowering drugs to a placebo control in men who had survived a recent heart attack (myocardial infarction). In this study, all patients also received the best medical care known at that time. Eligible men were randomized to receive either one of the five drugs or a placebo. They were followed for the recurrence of a major cardiovascular event such as death or a second heart attack. Many of the operational principles developed for this trial are still in use. Shortly after the CDP began, the NHLBI initiated several other large clinical trials evaluating modifications of major cardiovascular risk factors such as blood pressure in the Hypertension Detection and Follow-up Program (HDFP Cooperative Group 1982), cholesterol in the Coronary Primary Prevention Trial (The Lipid Research Clinics Program 1979) and simultaneous reduction of blood pressure, cholesterol, and smoking in the Multiple Risk Factor Intervention Trial (Domanski et al. 2002). These trials, all initiated within a short period of time, established the clinical trial as an important tool in the development of treatments for cardiovascular diseases. During this same period, the NHLBI launched trials studying treatments for blood and lung diseases. The methods used in the cardiovascular trials were applied to these trials as well.

In 1973, the National Eye Institute (NEI) also began a landmark clinical trial, the Diabetic Retinopathy Study (DRS) (Diabetic Retinopathy Study Research Group 1976). Diabetes is a risk factor for several organ systems diseases including cardiovascular and eye diseases. Diabetes causes progressive stages of retinopathy (damage to the retina of the eye), ultimately leading to severe visual loss or blindness. This trial evaluated a new treatment of photocoagulation by means of a laser device. Many of the concepts of the CDP were brought to the DRS by NIH statistical staff. Several other trials were launched by the NEI using the principles established in the DRS (e.g., the Early Treatment Diabetic Retinopathy Study (Cusick et al. 2005)).

Other institutes such as the National Cancer Institute (NCI) of the NIH

[4] The Coronary Drug Project Research Group (1975)

aggressively used the clinical trial to evaluate new treatments. The NCI established several clinical trial networks, or cancer cooperative groups, organized by either geographic regions (e.g., the Eastern Cooperative Oncology Group, or ECOG, the South Western Oncology Group, or SWOG), disease areas (e.g., the Pediatric Oncology Group, or POG), or treatment modality (e.g., Radiation Treatment Oncology Group, or RTOG). By 1990, most disease areas were using clinical trials to evaluate new interventions. Perhaps, the most recent development was in the AIDS Clinical Trial Group (ACTG) which was rapidly formed in the late 1980s to evaluate new treatments to address a rapidly emerging epidemic of Acquired Immune Deficiency Syndrome (AIDS) (DeMets et al. 1995). Many of the fundamental principles of trial design and conduct developed in the preceding two decades were reexamined and at times challenged by scientific, medical, patient, and political interest groups. Needless to say, these principles withstood the scrutiny and challenge.

Most of the trials we have mentioned were sponsored by the NIH in the U.S. or the Medical Research Council (MRC) in the U.K. Industry-sponsored clinical trials, especially those investigating pharmaceutical agents, evolved during the same period of time. Large industry-sponsored phase III outcome trials were infrequent, however, until the late 1980s and early 1990s. Prior to 1990, most industry-sponsored trials were small dose-finding trials or trials evaluating a physiological or pharmacology outcome. Occasionally, trials were conducted and sponsored by industry with collaboration from academia. The the Anturane Reinfarction Trial (The Anturane Reinfarction Trial Research Group 1978) was one such trial comparing a platelet active drug, sulfinpyrazone (anturane), to placebo in men following a heart attack. Mortality and cause-specific mortality were the major outcome measures. By 1990 many clinical trials in cardiology, for example, were being sponsored and conducted by the pharmaceutical industry. By 2000, the pharmaceutical industry was spending $2.5 billion dollars on clinical trials compared to $1.5 billion by the NIH. In addition, standards for the evaluation of medical devices as well as medical procedures are increasingly requiring clinical trials as a component in the assessment of effectiveness and safety.

Thus, the clinical trial has been the primary tool for the evaluation of a new drug, biologic, device, procedure, nutritional supplement, or behavioral modification. The success of the trial in providing an unbiased and efficient evaluation depends on fundamental statistical principles that we shall discuss in this and following chapters. The development of statistical methods for clinical trials has been a major research activity for biostatisticians. This text provides an introduction to these statistical methods but is by no means comprehensive.

Some of the most basic principles now used in clinical trial design and analysis can be traced to earlier research efforts. For example, an unplanned natural experiment to examine the effect of lemon juice on scurvy for sailors was conducted by Lancaster in 1600 as a captain of a ship for the East Indian Shipping Company (Bull 1959). The sailors on the ships with lemons on board

were free of scurvy in contrast to those on the other ships without lemons. In 1721, a smallpox experiment was planned and conducted. Smallpox was an epidemic that caused suffering and death. The sentences of inmates at the Newgate prison in Great Britain were commuted if they volunteered for inoculation. All of those inoculated remained free of smallpox. (We note that this experiment could not have been conducted today on ethical grounds.) In 1747, Lind (1753) conducted a planned experiment on the treatment of scurvy with a concurrent control group while on board ship. Of 12 patients with scurvy, ten patients were given five different treatments, two patients per treatment, and the other two served as a control with no treatment. The two sailors given fruit (lemons and oranges) recovered. In 1834, Louis (1834) described the process of keeping track of outcomes for clinical studies of treatment effect, and the need to take into consideration the patients' circumstances (i.e., risk factors) and the natural history of the disease.

While Fisher (1926) introduced the concept of randomization for agricultural experiments, randomization was first used for clinical research in 1931 by Amberson Jr., McMahon, and Pinner (1931) to study treatments for tuberculosis. As already described, Bradford Hill used randomization in the 1948 MRC tuberculosis trial (Hill 1971).

1.2 Ethics of Clinical Research

Clinical research in general and clinical trials in particular must be conducted in a manner that meets current ethical standards. Ethical standards change over time and can vary by geographical regions, societies, and even between individuals making the formulation of ethical standards complex and challenging. The ethical imperative for the establishment of ethical standards was starkly demonstrated by the discovery of Nazi atrocities carried out using concentration camp prisoners during World War II. As a result, the Nuremburg Code was established in 1947 and set standards for physicians and scientists conducting medical research (U.S. Government 1949). Two of the main tenants of the Nuremburg Code (summarized by Table 1.1) were that medical research must have patient consent and all unnecessary physical and mental suffering and injury should be avoided. The degree of risk should not exceed the potential benefit and a volunteer should be able to stop whenever they choose. The Declaration of Helsinki, first set forth in 1964 with later revisions (World Medical Association 1989), gives further guidance on the conduct of human research with specific reference to informed consent. The Belmont Report, summarized in Table 1.2, was issued in 1979 by the NIH as a guide establishing the need for *respect for persons*, especially those with diminished autonomy such as children and prisoners, the concept of *beneficence* to maximize the benefits while minimizing the risks, and the need for *justice* in the distribution of new experimental treatments.[5] The U.S. Department of Health

[5] National Commission for the Protection of Human Subjects of Biomedical and Behavioral Research (1979)

Table 1.1 *Nuremberg Code Principles**.

1. Voluntary consent
2. Experiment to yield results for good of society
3. Experiment based on current knowledge
4. Experiment to avoid all unnecessary suffering
5. No *a priori* reason to expect death
6. Risk not exceed importance of problem
7. Protect against remote injury possibilities
8. Conduct by scientifically qualified persons
9. Subject free to end experiment at any time
10. Scientist free to end experiment

*U.S. DHHS Institutional Review Board Guidebook Appendix 6 The Nuremberg Code, Declaration of Helsinki, and Belmont Report. www.hhs.gov/ohrp/irb/irb_appendices.htm

Table 1.2 *Principles established in the Belmont Report.*

1. Boundaries between practice and research
2. Basic ethical principles

   Respect for persons: recognition of the personal dignity and autonomy (i.e., self governance) of individuals and special protection of those with diminished autonomy (e.g., children, prisoners)

   Beneficence: obligation to protect persons from harm by maximizing potential benefits and minimizing potential risks of harm

   Justice: benefits and burdens of research be distributed fairly

3. Applications (parallels each basic ethical principle)

   Application of respect for persons: informed consent that contains information, comprehension, and voluntariness

   Application of beneficence: risk/benefit assessment is carefully considered in study design and implementation

   Application of justice: selection of research subjects must be the result of fair selection procedures

and Human Services (DHHS) through both the NIH[6] and the U.S. Food and Drug Administration (FDA)[7] also provide clinical research guidelines. In addition, NIH and FDA guidelines for monitoring trials will be discussed in detail in Chapter 10.

These federal documents discuss issues related to the experimental design, data management, and data analysis. The experimental design must be sound and not expose subjects to unnecessary risks while providing an adequate and fair test of the new experimental treatment. Studies must be sufficiently large to ensure reliable conclusions and the outcome assessed in an unbiased manner. Furthermore, adequate provisions must be made to protect the privacy of patients and the confidentiality of the data collected. The research plan must include provisions for monitoring of the data collected to ensure patient safety and to avoid prolonging an experiment beyond what is necessary to assess safety and effectiveness. All institutions that conduct federally sponsored research or research that is under federal regulation must have a body that reviews all proposed research to be conducted in their facility. These bodies are typically called Human Subjects Committees (HSC) or Institutional Review Boards (IRB). IRBs must comply with federal regulations or guidance documents as well as the local guidelines and ethical standards. An institution that fails to comply with these federal mandates may be sanctioned and have all federal funds for research terminated or put on hold until compliance is established. In addition, all federally regulated trials, including those sponsored by pharmaceutical companies and medical device companies, must comply with these IRB regulations. Thus these regulations, including those relevant to statistical design, data management, data monitoring, and data analysis, must be adhered to.

One of the key aspects of research studies in humans is the requirement for informed consent. Trial participants must be fully informed about the nature of the research, the goals, the potential benefits, and possible risks. The basic elements of the informed consent are given in Table 1.3. Participants must know that there may be alternatives to the treatment options in the study, and that their participation is entirely voluntary. Furthermore, even if they decide to start the trial, they may stop participation at any time. These issues have implications for the design, monitoring, and analysis of clinical trials discussed throughout this book.

Ethical concerns do not end with the local IRB. Journals that publish results of clinical trials are also establishing ethical standards. For example, the New England Journal of Medicine (Angell 1997) stated that they "will not publish unethical research regardless of scientific merit ... [T]he approval of the IRB in and informed consent of the research subjects are necessary but not sufficient conditions." One area of dispute is the conduct of clinical trials in developing countries in which many citizens do not have access to mod-

[6] http://grants.nih.gov/grants/guide/notice-files/not98-084.html
[7] http://www.fda.gov/cber/guidelines.htm

Table 1.3 *Eight basic elements of informed consent (45 CFR 46.116).*

1. A statement that the study involves research, an explanation of the purpose(s) of the research, the expected duration of the subject's participation, and a description of the research procedures (e.g., interview, observation, survey research).
2. A description of any reasonably foreseeable risks or discomforts for the subjects. Risks should be explained to subjects in language they can understand and be related to everyday life.
3. A description of any benefits to the subject and/or to others that may reasonably be expected from the research.
4. A disclosure of appropriate alternative procedures or courses of treatment, if any, that might be advantageous to the subject.
5. A statement describing the extent, if any, to which the confidentiality of records identifying the subject will be maintained.
6. For research involving more than minimal risk, a statement whether compensation is available if injury occurs and, if it is, what it consists of and from whom further information may be obtained.
7. An explanation of whom to contact for answers to pertinent questions about the research and research subject's rights. The name and phone number of the responsible faculty member as well as contact information for an IRB must be included for these purposes. In addition, if the project involves student research, the name and phone number of the student's adviser/mentor must also be included.
8. A statement that research participation is voluntary and the subject may withdraw from participation at any time without penalty or loss of benefits to which the subject is otherwise entitled. If the subject is a patient or client receiving medical, psychological, counseling, or other treatment services, there should be a statement that withdrawal will not jeopardize or affect any treatment or services the subject is currently receiving or may receive in the future. If the subject is a prisoner, there should be a statement that participation or non-participation in the research will have no effect on the subject's current or future status in the prison. If a survey instrument or interview questions are used and some questions deal with sensitive issues (including but not limited to illegal behavior, mental status, sexuality, or sexual abuse, drug use, or alcohol use) the subjects should be told they may refuse to answer individual questions.

ern western medicine. The control therapies used in trials must be consistent with the best that particular country can afford or deliver, yet these therapies are likely to be inferior to the standard of care in the United States or Europe. If such trials cannot be performed, the medical treatments in these countries will not advance through direct and rigorous evaluation of alternative treatments, perhaps those that are more affordable or more practical. Love and Fost (2003) comment on a trial in Vietnam that compared a simple and affordable treatment for breast cancer to the local standard of care, even though the new treatment is viewed as far inferior to that provided to women in the United States. In fact, Love and Fost (1997) report that this simple therapy, that involves removing the patient's ovaries and providing tamoxifen, a drug affordable by these patients, was clinically superior to the local traditional standard of care. Thus, this trial provided a substantial advance for the women of Vietnam while answering fundamental scientific questions. The ethical issues are non-trivial, however, and must be given careful consideration. Regardless of the outcome of this debate, it is clear that the standard for statistical conduct in all trials must be the highest possible in order to meet ethical criteria, a responsibility borne largely by the trial biostatistician.

1.3 Types of Research Design and Types of Trials

Medical research makes progress using a variety of research designs and each contributes to the base of knowledge regardless of their limitations. The most common types of clinical research designs are summarized in Table 1.4. The simplest type, and which is often used, is the case report or anecdote—a physician or scientist makes an astute observation of a single event or a single patient and gains insight into the nature or the cause of a disease. An example might be the observation that the interaction of two drugs causes a life threatening toxicity. It is often difficult, however, to distinguish the effects of a treatment from those of the natural history of the disease or many other confounding factors. Nevertheless, this unplanned anecdotal observation remains a useful tool. A particularly important example is a case report that linked a weight reduction drug with the presence of heart valve problems (Mark et al. 1997).

Epidemiologists seek associations between possible causes or risk factors and disease. This process is necessary if new therapies are to be developed. To this end, observational studies are typically conducted using a larger number of individuals than in the small case report series. Identifying potential risk factors through observational studies can be challenging, however, and the scope of such studies is necessarily limited (Taubes 1995).

Observational studies can be grouped roughly into three categories (Table 1.4), referred to as *retrospective*, *cross-sectional*, and *prospective*. A *case-control* study is a retrospective study in which the researcher collects retrospective information on *cases*, individuals with a disease, and *controls*, individuals without the disease. For example, the association between lung cancer

Table 1.4 *Types of research.*

Case Report An astute clinician or scientist observes an event or situation indicating a potential problem that is unusual or never before noted.

Observational A class of studies that are characterized by data collection on a cohort of individuals with the intent of correlating potential risk factors with clinical outcomes.

- **Retrospective** This design observes individuals or cases who have an event or disease diagnosis and then collects data on prior medical history or exposure to environmental, behavior, and other factors. A control group without the event or disease is also identified. The goal is to identify specific exposures that are more frequent in cases than control individuals.
- **Cross-Sectional** A cohort of individuals is observed and data collected at a single point in time. The cohort will have a mixture of individuals with disease and without. The goal is to find associations between exposure and the presence of the disease.
- **Prospective** A cohort of individuals is identified and followed prospectively, or forward in time. Exposure variables measured at the beginning are correlated with incident or new events. The goal is to identify disease risk factors.

Clinical Trial An experiment in which a group of individuals is given an intervention and subsequent outcome measures are taken. Results of intervention are compared to individuals not given the intervention. Selection of control group is a key issue.

- **Historical** Historical controls are obtained by using data from previous individuals not given the experimental intervention.
- **Concurrent** Data from individuals who are not being given the intervention are collected during the same period of time as from the intervention group.
- **Randomized** The assignment of intervention or control to individuals is through a randomization process.

Table 1.5 *Research biases.*

Selection Bias Bias affecting the interventions that a patient may receive or which individuals are entered into the study.
Publication Bias Studies that have significant (e.g., $p < 0.05$) results are more likely to be published than those that are not. Thus, knowledge of literature results gives a biased view of the effect of an intervention.
Recall Bias Individuals in a retrospective study are asked to recall prior behavior and exposure. Their memory may be more acute after having been diagnosed with a disease than the control individuals who do not have the disease.
Ascertainment Bias Bias that comes from a process where one group of individuals (e.g., intervention group) is measured more frequently or carefully than the other group (e.g., control).

and smoking was established primarily through a number of large case-control studies (Shopland (ed) 1982). A large number of patients with lung cancer were interviewed and information was obtained on their medical history and behavior. The same information was obtained on individuals not having the diagnosis of lung cancer. Comparisons are typically made between the frequency of factors in medical history or lifestyle. In the case of lung cancer, it became apparent, and overwhelmingly convincing, that there was a substantially higher frequency of a history of smoking in those individuals who developed lung cancer compared to those individuals free of lung cancer. The case-control design has proven to be quite useful, especially for relatively rare diseases such as lung cancer. As with all observational studies, however, the case control design has limitations and is vulnerable to bias. For example, associations do not imply causation, but rather that the proposed risk factor and the disease occur together, either by chance or because of a third, possibly unknown, factor. The choice of a control group is also critical and bias can be introduced if it is not chosen properly. Control groups selected from the literature or from previous cohorts are subject to publication bias or selection bias. Furthermore, both case and control groups are vulnerable to recall bias (see Table 1.5).

The *cross-sectional* design compares individuals in a defined population at a particular moment in time, again looking for associations between potential risk factors and disease frequency. In this design, some of the biases present in the case control design can be minimized or eliminated. Publication bias and recall bias are eliminated. This design, however, can only evaluate those who are alive at the time of the evaluation and therefore a degree of bias is inherent. Again, associations that are identified are not necessarily causative

factors, but rather factors that coexist with the disease under investigation. Examples of cross-sectional studies are the Wisconsin Epidemiologic Study of Diabetic Retinopathy (WESDR) (Klein et al. 1985) and the Beaver Dam Eye Study (Klein et al. 1991). These studies initially were established as cross sectional studies to identify possible risk factors for eye diseases. WESDR, for example, identified serum levels of glycosylated hemoglobin as a risk factor for the incidence and progression of diabetic retinopathy.

A third observational design is the *prospective cohort* study in which a cohort of individuals is identified and followed forward in time, observing either the incidence of various diseases of interest or survival. At the beginning, extensive information is collected on individuals including, for example, medical history, blood and urine chemistry, physiologic measurements, and perhaps genetic material. Associations are sought between all of these baseline data and the occurrence of the disease. One of the earliest and best known prospective studies was the Framingham Heart Study (FHS) (Dawber et al. 1951). Several hundred individuals in the town of Framingham, Massachusetts were identified in 1950 and followed for the next three decades. Initially, a large amount of information based on medical history and physical examination was collected. The FHS was primarily interested in heart disease and from this study, researchers identified high cholesterol and other elevated lipids, high blood pressure, smoking, and diabetes as possible risk factors for heart disease. Again, these associations did not establish causation. Clinical trials conducted later, and described later, examined the impact of modification of these possible risk factors and the reduction of disease incidence. Nonetheless, the FHS was essential in the identification of these risk factors and played a landmark role.

Because of the potential for bias, observational studies have led to many false positive associations (Taubes 1995) that could not be replicated. Examples include the association of high cholesterol and rectal cancer, smoking and breast cancer, vasectomy and prostate cancer, red meat and either breast or colon cancer, and excessive water consumption and bladder cancer. Despite these false positive associations, the observational design is an important component of the research cycle. While replication of results are an essential to ensure credibility of the results, this may not be sufficient. For example, observational studies have suggested that low serum beta-carotene is associated with an increase in lung cancer. A synthetic beta-carotene tablet was developed and three trials were conducted to test whether increasing the level of serum beta-carotene through dietary supplementation resulted in a lower incidence of either lung cancer, or cancer in general. The Alpha-Tocopheral Beta-Carotene trial (ATBC)[8] was a trial in a cohort of Finnish male heavy smokers. Contrary to the observational studies, the incidence of lung cancer increased in those given beta-carotene supplements despite the documented increase in serum beta-carotene. This result was repeated in a U.S. based trial

[8] The Alpha-Tocopherol, Beta-Carotene Cancer Prevention Study Group (1994)

in smokers and asbestos workers, referred to as CARET (Omenn et al. 1994). A third trial, the Physicians Health Study (PHS) (Hennekens et al. 1996), evaluated beta-carotene supplementation in a cohort of U.S. male physicians. Again, while serum beta-carotene was increased, the incidence of lung cancer and all cancers did not change. Remarkably, the observation of an association between the baseline level of beta-carotene and the incidence of lung cancer was found in all three trials, confirming the previous observational studies. Still, increasing the level with a beta-carotene supplement did not reduce the incidence of lung cancer. Replication did not guarantee that the association is causal.

Thus, whether case-control studies, cross-sectional studies, or prospective studies are used, the role of the observational studies is critical in identifying factors to be further studied as possible risk factors for disease progression or occurrence. The next step is to attempt to modify the proposed risk factor and determine if the incidence can be lowered. Laboratory research and small clinical studies are conducted to identify safe and effective drugs, biologics, or devices that can modify the risk factor. This process can take months or years. For cholesterol and blood pressure, drugs were eventually developed that modified blood pressure and cholesterol levels. Later, trials using these drugs established that changes in these risk factors resulted in a reduction in heart disease progression and death. For example, the Hypertension Detection Follow-up Program (HDFP) (HDFP Cooperative Group 1982) demonstrated the positive benefits of lowering blood pressure in patients with mild to moderate hypertension. The Scandinavian Study of Simvistatin (Scandanavian Simvistatin Survival Study 1994) established that treatments that lower cholesterol may also lower the risk of death and heart attacks. This text is focused on the statistical issues in the design, conduct, and analysis of such trials. These trials are essential in the completion of the research process.

In fact, the research process is a dynamic interaction between observation, laboratory results, and clinical trials illustrated by Figure 1.1. All three elements are essential and may be conducted simultaneously as researchers probe all aspects of a medical problem.

Clinical trials are categorized into 4 phases, summarized in Table 1.6. These clinical trial phases will be described in more detail in Chapter 3. Briefly, although the precise goals and designs may vary between disease areas, the goal in phase I is usually to determine the maximum dose that can be tolerated without excessive adverse effects. Typically, phase I trials are conducted either in healthy volunteers or in patients who have failed all regular treatments. phase II trials are usually conducted to evaluate the biological activity of the new drug to determine if it evokes the response that was expected and warrants further development. Phase III trials are comparative trials that evaluate the effectiveness of the new treatment relative to the current standard of care. These trials may add the new treatment to the standard of care and compare that to the standard of care alone. Some trials compare two known active agents to determine which is superior, or in some cases, to determine if the

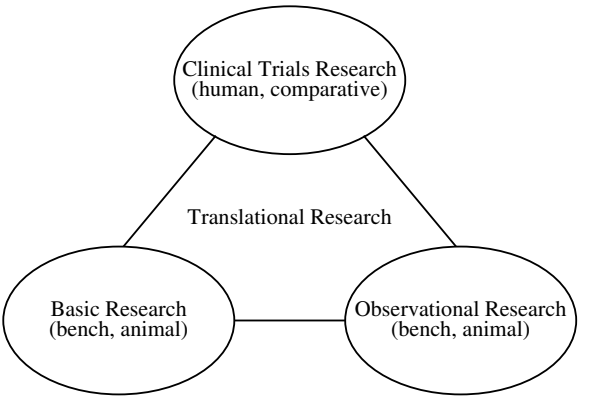


Figure 1.1 *The research triangle.*

Table 1.6 *Clinical trial phases.*

Preclinical	Once a risk factor is identified, laboratory research is conducted to identify a means to modify the risk factor, testing it in the laboratory and often in animal models.
Phase I	With a new intervention available from laboratory research, the first step is to determine if the intervention can be given to humans, by what method, and in what dose.
Phase II	Trials in the second phase typically measure how active the new intervention is, and learn more about side effects.
Phase III	Trials in the third phase compare whether the new intervention is more effective than a standard control intervention.

two have similar effectiveness. Phase IV trials usually follow patients who have completed phase III trials to determine if there are long term adverse consequences.

Phase III trials are also classified according to the process by which a control arm is selected. *Randomized control trials* assign patients to either the new treatment or the standard by a randomization method, described in Chapter 5. Non-randomized phase III trials can be of two general types. The *historical control* trial compares a group of patients treated with the new drug or device to a group of patients previously treated with the current standard of care. A *concurrent control* trial, by contrast, compares patients treated with the new treatment to another group of patients treated in the standard manner at the same time, for example, those treated at another medical facility or clinic. As will be discussed in Chapter 5, the randomized control trial is considered to be

the gold standard, minimizing or controlling for many of the biases to which other designs are subject. Trials may be single center or multiple center, and many phase III trials are now multinational.

Trials may also be classified by the nature of the disease process the experimental intervention is addressing. Screening trials are used to assess whether screening individuals to identify those at high risk for a disease is beneficial, taking into account the expense and efforts of the screening process. For example, a large cancer screening trial is evaluating the benefits of screening for prostate, lung, colon, and ovarian cancer (Prorok et al. 2000). These trials must, by nature, be long term to ascertain disease incidence in the screened and unscreened populations. Screening trials are conducted under the belief that there is a beneficial intervention available to at-risk individuals once they are identified. Primary prevention trials assess whether an intervention strategy in a relatively healthy but at risk population can reduce the incidence to the disease. Secondary prevention trials are designed to determine whether a new intervention reduces the recurrence of the disease in a cohort that has already been diagnosed with the disease or has experienced an event (e.g., heart attack). Therapeutic or acute trials are designed to evaluate an intervention in a patient population where the disease is acute or life threatening. An example would be a trial that uses a new drug or device that may improve the function of a heart that has serious irregular rhythms.

1.4 The Need for Clinical Trials

Since the introduction of the modern clinical trial in 1950, a great deal of medical research has been conducted into the causes and possible treatments of numerous diseases. Diagnostic methods have improved so that disease or disease progression can be detected earlier and more accurately. During this time, the number of approaches to the treatment or prevention of disease has increased dramatically and these must be evaluated to establish their effectiveness and proper role. Since the clinical course of many diseases is complicated, determining if a new therapy is effective or superior to existing treatments is not an easy task. It often requires a systematic evaluation using large numbers of patients and astute clinical observation.

The clinical trial has become a standard tool because it is the most definitive and efficient method for determining if a new treatment is more effective than the current standard or has any effectiveness at all. Observational studies have potential for bias and one cannot conclusively determine from an uncontrolled study whether differences in outcomes (or lack of differences) can be directly attributed to the new intervention. As discussed previously, many potential risk factors have been identified through uncontrolled studies and later shown to be spurious (Taubes 1995).

Controlled clinical trials are also an effective mechanism to distinguish incidence of side effects and adverse effects due to the therapy from those caused by the disease process itself. For example, in the Coronary Drug Project,

cardiac arrhythmias were observed in 33% of the patients on either Niacin or Clofibrate, two of the drugs being tested.[9] On the other hand, 38% of the patients on the placebo arm had cardiac arrhythmias as well. Without the control arm, one might have associated the adverse effect with the drugs instead of recognizing that it is a consequence of the underlying disease. As another example, 7.5% of the patients on clofibrate experienced nausea, but 6.2% of the placebo patients did as well. Again, the nausea is not attributable to the drug, but this might not have been realized without the control comparison.

One of the most compelling reasons for the use of clinical trials is that if they are not conducted, new, but ineffective or even harmful, treatments or interventions can become part of medical practice. Many ineffective or harmful interventions have been in use for a long time before their effects were understood. One of the classic examples is the use of high dose oxygen in infants born prematurely.

Children born prematurely typically have lungs that are not fully developed and thus have difficulty breathing. As described by Silverman (1979), the practice of giving premature infants high doses of oxygen began in the 1940s and eventually became established as the standard of care. During the same time period, an epidemic of retrolental fibroplasia, which often leads to blindness in premature infants, began. A careful review of case records indicated that these affected premature infants received the "state of the art" medical care, including high dose oxygen. In the early 1950s, some researchers began to suspect the high dose oxygen but the evidence was not convincing. Furthermore, a careful study of the use of oxygen was ethically challenging since this was the accepted standard of care. One trial attempted to examine this question by randomizing premature infants to receive either high (standard) or low dose oxygen. Because of the belief that high dose was the ethical treatment, nurses turned up the oxygen levels at night in those infants randomized to the low dose oxygen group. Later, in 1953, another randomized clinical trial was launched in 800 premature infants. Results indicated that 23% of the infants receiving the high dose oxygen were blinded compared to 7% in those receiving a low dose (50% of standard) oxygen when needed. This trial confirmed earlier suspicions and, when the results were published in 1954, the practice diminished. It was estimated that perhaps 10,000 infants had been blinded by the practice of high dose oxygen administration. A widely used but untested intervention was ultimately shown to be harmful.

The high dose oxygen story is not the only case of untested interventions being ineffective or harmful but in widespread use. Many common, accepted treatments have never been formally tested. The FDA regulatory laws were not in effect prior to 1968 so that drugs developed prior to that time were "grandfathered." Medical devices are regulated by the FDA as well but as a result of different legislation having different requirements. Many devices may have been tested to assess functionality but not necessarily clinical effective-

[9] The Coronary Drug Project Research Group (1975)

ness. Surgical and other procedures do not fall under FDA regulation and thus many have not been rigorously tested. The same is true of many behavioral modifications or nutritional supplements.

The Intermittent Positive Pressure Breathing (IPPB) trial[10] is an example in which a device used for patients with advanced pulmonary obstructive disease became an established, expensive practice but was later shown to have no clinical benefit. The IPPB delivered bronchodilator drugs deep into the lung under pressure based on the hypothesis that distributing the drug throughout the entire lung would be beneficial. The treatment using IPPB requires an expensive device and technical staff trained in the use of this device. When IPPB was compared to a inexpensive hand held nebulizer that also delivered the drug, the clinical effect was the same, as measured by standard pulmonary function tests. Over time, the use of this expensive but ineffective therapy has diminished.

The treatment of cardiac arrhythmias provides another convincing example. Cardiac arrhythmias are associated with a higher incidence of sudden death and drugs developed to suppress arrhythmias were approved by the FDA for use in high risk patients. Cardiologists, however, began using these drugs more broadly in lower risk patients. The Cardiac Arrhythmia Suppression Trial (CAST)[11] was designed to test whether the use of these new drugs would in fact reduce the risk of sudden death and total mortality. CAST was a well designed, randomized placebo controlled trial. Shortly after the trial began, when approximately 15% of the mortality information had accrued, the trial was terminated with a statistically significant increase in both sudden death and total mortality among subjects receiving anti-arrhythmic agents. A class of drugs that was rapidly becoming part of medical practice was discovered to be harmful to patients with cardiac arrhythmias.

Finally, the use of an effective intervention can be delayed because trials were not conducted in a timely fashion. Chronic heart failure (CHF) is a disease of a failing heart that is not able to pump efficiently, and the risk of mortality increases with the progression of the disease. The efficiency of the heart is measured by the ejection fraction—how much of the heart chamber is emptied with contraction relative to when it has been filled. A class of drugs known as beta-blockers were known to be effective in lowering blood pressure in individuals with high blood pressure and in slowing or controlling the heart rhythm following a heart attack. Since CHF patients already are having trouble with an inefficient heart, treating them with a drug that would slow down the heart rhythm and lower blood pressure seemed like the wrong approach. For several years, using beta-blockers in CHF patients was discouraged or proscribed even though there was research suggesting that beta-blockers may in fact be beneficial to CHF patients. Three trials were conducted in CHF

[10] The Intermittent Positive Pressure Breathing Trial Group (1983)
[11] The Cardiac Arrhythmia Suppression Trial (CAST) Investigators (1989)

patients, using different beta-blocking drugs, and all three demonstrated significant and substantial reduction in mortality, contrary to common belief.[12]

While we cannot afford to study every new medical intervention with a carefully controlled clinical trial, we must study those for which the outcome is serious and with the potential for serious adverse effects. Regulatory agencies world-wide require that most new drugs and biologics must be rigorously tested. Devices are increasingly being tested in a similar manner. It not clear, however, when the new use of an existing drug or device must receive the same rigorous evaluation. Regulatory requirements address many but not all of these circumstances. No requirements exist for procedures and behavioral modifications. Thus, society and the medical profession together must make judgments, realizing that a degree of risk is assumed when an intervention is not tested by a clinical trial. In some cases, that risk may be justified, but those circumstances should be rare.

1.5 The Randomization Principle

As discussed in Chapters 2, 3, and 5, the process of randomization has three primary benefits. The first is that randomization guarantees that assigned treatment is stochastically independent of outcome. In non-controlled studies, confounding occurs when *exposure* to the agent of interest is associated with or the result of the disease in question, often resulting in spurious conclusions regarding causality. Randomization ensures that this cannot happen—any observed association is either causal or the result of chance, and the latter can be well controlled. Second, randomization tends to produce comparable groups with regard to measured and unmeasured risk factors, thus making the comparison between the experimental and standard or control groups more credible (Byar et al. 1976). Finally, randomization justifies the analysis typically conducted without depending on external distribution assumptions. That is, common tests such as t-tests and chi-square tests are approximations to the randomization test associated with the randomization procedure. This principle has been discussed generally by Kempthorne (1977), for example, and specifically for clinical trials by Lachin (1988b). This principle will be discussed in more detail in Chapter 5.

1.6 Timing of a Clinical Trial

Often there is a relative narrow window of time during which a trial can be conducted. As previously indicated, if an intervention becomes part of the established standard of care without rigorous safety and efficacy assessments, it can become ethically challenging to rigorously test its safety and effectiveness. Thus, it is important to evaluate a new intervention before it becomes part of clinical practice. We have already discussed several examples where

[12] The International Steering Committee on Behalf of the MERIT-HF Study Group (1997), Krum et al. (2006), Packer et al. (2001)

interventions become part of practice before a trial has been conducted. These examples include the use of high dose oxygen in premature infants (Silverman 1979), intermittent positive pressure breathing device in chronic obstructive pulmonary patients[13] and a class of arrhythmia drugs in patients with cardiac arrhythmias.[14] Other examples include the use of hormone replacement therapy (HRT) for reducing the risk of heart disease in post menopausal women,[15] and the use of coronary artery bypass graft (CABG) surgery to treat patients who were experiencing symptoms such as angina (heart pain) due to the occlusion (narrowing) of coronary vessels from atherosclerosis (Healy et al. 1989).

CABG is a surgical procedure that takes healthy vessels from other parts of the body and grafts them onto the heart to bypass the occluded segments of the coronary vessels. CABG became a rapidly accepted surgery procedure before being evaluated in randomized clinical trials. When the Coronary Artery Surgery Study (CASS) was conducted, there was a reluctance on the part of many cardiac surgeons to randomize their patients to either medical therapy or CABG. As a result, a registry of patients who were to undergo CABG was a part of CASS (CASS Principal Investigators and their Associates 1983). Many more patients were entered into the registry than were entered into the randomized trial. The randomized portion of CASS demonstrated that CABG did not reduce mortality relative to medical therapy in the less advanced cases—those with fewer occluded coronary vessels—in spite of its increasingly widespread use.

Conversely, large phase III confirmatory trials should be designed and conducted only after sufficient background information regarding the population and the new intervention is available. Information about the level of risk or disease incidence in the population of interest is required before the entry criteria and sample size can be determined. Researchers also must have knowledge about the safety of the intervention, the dosing schedule, and the stability of the treatment formulation. In the trial design, the clinical outcomes of interest must be determined as well as the expected size of the effect of the intervention. Financial resources and patient availability must also be determined.

Launching a trial prematurely, before adequate knowledge is available for proper design, may lead to operational problems. For example, researchers may find the entry criteria too strict and the number of patients eligible is much less than required, making recruitment goals unattainable, or, the risk level of the trial population may be less than expected resulting in recruitment goals that are too small. If insufficient information is available, the trial may have to be suspended until the design can be corrected, or if that is not possible, the trial may have to be terminated wasting time and resources before the trial can begin again or another trial designed.

[13] The Intermittent Positive Pressure Breathing Trial Group (1983)
[14] The Cardiac Arrhythmia Suppression Trial (CAST) Investigators (1989)
[15] Writing Group for the Women's Health Initiative Randomized Controlled Trial (2002)

1.7 Trial Organization

In the mid 1960s, when the National Heart Institute was planning a series of risk factor intervention trials, beginning with the CDP, they planned for the organizational structure of such trials. A task force was commissioned, chaired by Dr Bernie Greenberg, that issued a 1967 report that became known as the Greenberg Report. This report was formally published in 1988, long after its impact on the early NIH trials (Heart Special Project Committee 1998). As

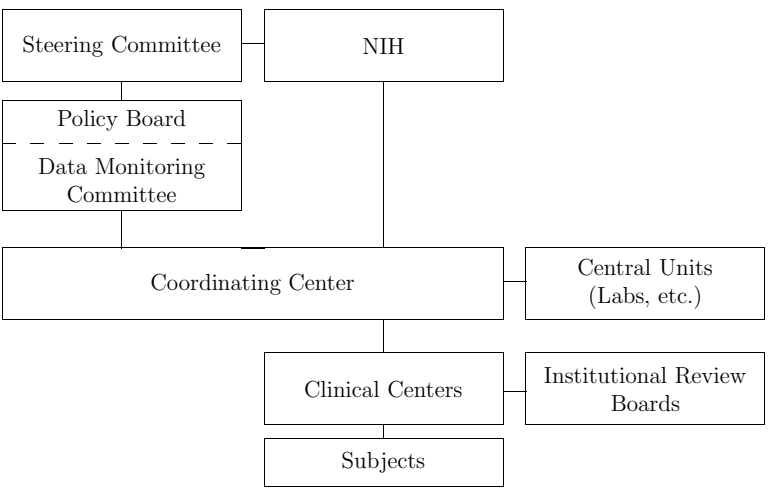


Figure 1.2 *NIH model. Reprinted with permission from Fisher, Roecker, and DeMets (2001). Copyright © 2001, Drug Information Association.*

shown in Figure 1.2, there are several key functional components. All trials must have a *sponsor* or *funding agency* to pay for the costs of the intervention, data collection, and analysis. Funding agencies often delegate the management of the trial to a *steering committee* or *executive committee*, a small group composed of individuals from the sponsor and the scientific investigators. The steering committee is responsible for providing scientific direction and to monitor the conduct of the trial. The steering committee may appoint working committees to focus on particular tasks such as recruitment, intervention details, compliance to intervention, outcome assessment, as well as analysis and publication plans. Steering committees usually have a chair who serves as the spokesperson for the trial. A network of investigators and clinics is typically needed to recruit patients, apply the intervention and other required patient care, and to assess patient outcomes. Clinical sites usually will have a small staff who dedicate a portion of their time to recruit patients, deliver the intervention, assess patient responses, and complete data collection forms. For some trials, one or more central laboratories are needed to measure blood chemistries in a uniform manner or to evaluate electrocardiograms, x-rays,

eye photographs, or tumor specimens. Figure 1.3 depicts a modification of the NIH clinical trial model that is often used for industry sponsored trials (Fisher et al. 2001). The major difference is that the data coordinating center operation depicted in Figure 1.2 has been divided into a data management center and a statistical analysis center. The data management center may be internal to the sponsor or contracted to an outside organization. The statistical analysis center may also be internal or contracted to an external group, often an academic-based biostatistics group. As described in Chapter 10, careful mon-

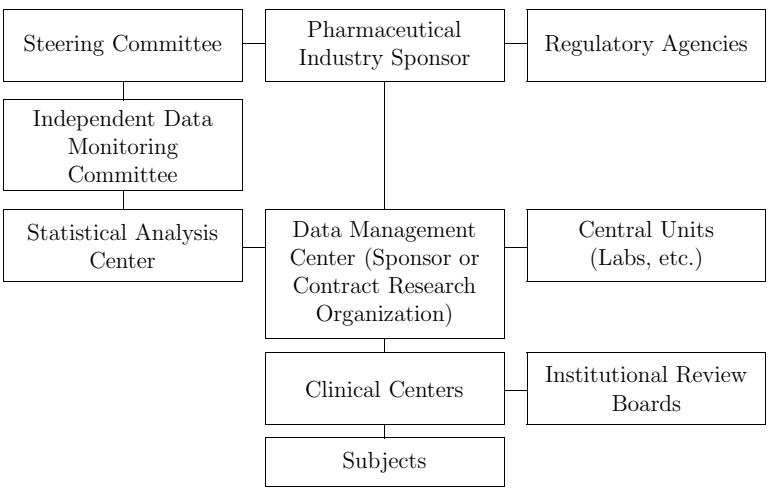


Figure 1.3 *Industry-modified NIH model. Reprinted with permission from Fisher, Roecker, and DeMets (2001). Copyright © 2001, Drug Information Association.*

itoring of trials is ethically and scientifically mandated. This responsibility is largely the task of the Data Monitoring Committee (DMC), also referred to as the Data and Safety Monitoring Board (DSMB). The DMC is made up of experts in the clinical condition of the patient or participant, the intervention being studied, epidemiology of the disease, statistics, and clinical trials. Their role is to monitor the accumulating data and to terminate a trial early if there is convincing evidence of harm, if the intervention has shown overwhelming benefit earlier than expected, or has no chance of being successfully completed. In some circumstances, the DMC may recommend modifications to the trial.

A statistical and data management center is also necessary and is where much of the day to day trial activity takes place. In the original NHLBI model, these two functions are performed in one center, referred to as a trial coordinating center, although these functions can be separated into two centers, a statistical analysis center and a data management center. These centers design and implement the data collection process, including processes for data entry, data editing, and data quality control. The statistical analysis of safety and