CLINICAL TRIALS - DESIGN, CONDUCT AND ANALYSIS (PART I)

CURTIS L. MEINERT

Monographs in Epidemiology and Biostatistics
edited by Abraham M. Lilienfeld

1 THE EPIDEMIOLOGY OF DEMENTIA
James A. Mortimer and Leonard M. Schuman
1981

2 CASE CONTROL STUDIES
Design, Conduct, Analysis
James J. Schlesselman
1982

3 EPIDEMIOLOGY OF MUSCULOSKELETAL DISORDERS
Jennifer L. Kelsey
1982

4 URBANIZATION AND CANCER MORTALITY
The United States Experience, 1950–1975
Michael R. Greenberg
1983

5 AN INTRODUCTION TO EPIDEMIOLOGIC METHODS
Harold A. Kahn
1983

6 THE LEUKEMIAS
Epidemiologic Aspects
Martha S. Linet
1984

7 SCREENING IN CHRONIC DISEASE
Alan S. Morrison
1985

8 CLINICAL TRIALS
Design, Conduct, and Analysis
Curtis L. Meinert
1986

9 VACCINATING AGAINST BRAIN DYSFUNCTION SYNDROMES
The Campaign Against Rubella and Measles
Ernest M. Gruenberg
1985

10 OBSERVATIONAL EPIDEMIOLOGIC STUDIES
Jennifer L. Kelsey, W. Douglas Thompson, Alfred S. Evans
1986

Monographs in Epidemiology and Biostatistics
Volume 8

CLINICAL TRIALS
Design, Conduct, and Analysis

Curtis L. Meinert, Ph.D.
Professor of Epidemiology and Biostatistics
School of Hygiene and Public Health
The Johns Hopkins University

In collaboration with
Susan Tonascia, M.Sc.
Research Associate
School of Hygiene and Public Health
The Johns Hopkins University

New York Oxford
OXFORD UNIVERSITY PRESS
1986

Oxford University Press
Oxford New York Toronto
Delhi Bombay Calcutta Madras Karachi
Petaling Jaya Singapore Hong Kong Tokyo
Nairobi Dar es Salaam Cape Town
Melbourne Auckland

and associated companies in
Beirut Berlin Ibadan Nicosia

Published by Oxford University Press, Inc.,
200 Madison Avenue, New York, New York 10016

Library of Congress Cataloging-in-Publication Data

Meinert, Curtis L.
Clinical trials.

(Monographs in epidemiology and biostatistics; v. 8)
Bibliography: p.
Includes indexes.
1. Clinical trials. I. Tonascia, Susan. II. Title.
III. Series. [DNLM: 1. Clinical Trials. W1 M0567Lt v.8
/ QV 771 M514c]
R853.C55M45 1986 610'.72 85-11530
ISBN 0-19-503568-2

Printing (last digit): 9 8 7 6 5 4 3 2 1

Printed in the United States of America
on acid-free paper

To
Susie, Julie, Nancy, and Jill

my wife and daughters for their help,
encouragement, forbearance, and understanding

Preface

And further, by these, my son, be admonished: Of making many books there is no end; and much study is a weariness of the flesh.
Ecclesiastes 12:12

This book consists of seven parts:

It is intended as a general reference for practitioners of clinical trials. The main focus is on trials involving uncrossed treatments and a clinical event as the outcome measure. It is not concerned with trials designed to assess bioavailability or with trials involving crossover designs. However, this is not to say that it is of no value for researchers with such interests, since some of the design and operating principles and practices described herein extend to such trials as well. Parts of this book, such as the chapters concerned with sample size calculation, randomization, forms design, quality assurance, and reporting procedures, apply to most kinds of trials.

This book deals with single-center as well as multicenter trials, as defined in Chapter 4. No distinction is made between the two types in most of the chapters, because the design and operating practices are largely the same for both. There are only two chapters, 5 and 23, that deal exclusively with multicenter trials, and even they have some relevance to single-center trials.

Appendix A contains a glossary of terms and acronyms used in this book and serves as a starting point for a dictionary of terms for clinical trials in general. Appendix B contains operational information for 14 of the trials referenced. Tabulations based on the information contained in this appendix appear in various chapters of the book. Appendix I contains a combined bibliography of references cited in the various chapters and appendixes (except B and C). References in the combined bibliography have been arranged alphabetically by first author and then chronologically. The reference lists in Table B-3, Appendix B, for the studies sketched, are in chronological order. Journal abbreviations used in the reference listings throughout correspond to those used by the National Library of Medicine in *Index Medicus* and MEDLINE. The other five appendixes relate to specific chapters in the book.

The impetus for this book emerged from a long-standing involvement in clinical trials, beginning with the University Group Diabetes Program in 1961. The urge to develop a general text concerned with the design and conduct of clinical trials led to development of an initial draft in the spring of 1972. The emphasis in that and subsequent drafts during the next two years focused exclusively on a few large-scale multicenter trials. Work continued, but at a decelerating rate, until it came to a virtual halt by 1975, primarily because of other work commitments. The work lay dormant until late 1978 when, while still at the University of Maryland School of Medicine, I was persuaded to start anew by the late Abraham Lilienfeld. The revised outline involved 8 chapters. It gradually expanded to the current size.

Writing proceeded slowly until my move to the Department of Epidemiology of The Johns Hopkins University School of Hygiene and Public Health in late 1979, where I was faced with the challenge of developing a course on the design and conduct of clinical trials. That teaching effort and Susan Tonascia's participation in that activity helped me to organize my thoughts and to collect the materials needed for this book. I am indebted to her for her help.

Baltimore, Maryland C.L.M.
November 1985

Acknowledgments

I wish to begin by expressing my thanks to two individuals who stimulated and focused my interest in clinical trials. My sincere thanks to Jacob Bearman for his confidence in me when I had none and to Chris Klimt for his role in developing and promoting my career and for giving me the chance to learn about clinical trials by doing. I am indebted as well to the late Abraham Lilienfeld for his words of encouragement and editorial help during this writing effort.

One other individual, James Tonascia, deserves special thanks. I have benefited immensely from his teaching and counsel. His ideas and input are reflected in various places throughout this book. In fact, his notes, developed for an advanced course in epidemiology at Johns Hopkins, provided the starting point for much of what is contained in Chapter 9 and for some of Chapters 10 and 18 as well. In addition, he has spent hours reviewing various parts of this book, including all of Chapter 9 and parts of Chapter 10 and the Glossary.

Various others helped on specific parts of the book. I wish to thank each of them for their contribution. They include:

- Helen Abbey, for review of Chapter 9
- Barbara Andreassen, for help on Chapter 12 and art work on figures in the book
- Steven F. Bingham, for review of the VACSP #43 sketch in Appendix B
- Thomas Blaszkowski, for review of glossary terms and for referencing help concerning administration of government grants and contracts
- Robert Bradley, for referencing help on Chapter 7
- Paul L. Canner, for review of the Coronary Drug Project sketch in Appendix B and for information concerning procedures used in the CDP
- Jeffrey A. Cutler, for review of the Multiple Risk Factor Intervention Trial sketch in Appendix B
- Marie Diener, for help on Chapter 7
- Lloyd Fisher, for review of the Coronary Artery Surgery Study sketch in Appendix B and for information concerning procedures used in CASS
- Lawrence Friedman, for review of glossary terms and for review of the Aspirin Myocardial Infarction Study sketch in Appendix B
- Curt Furberg, for review of glossary terms
- Barbara S. Hawkins, for review of glossary terms, for review of the Macular Photocoagulation Study sketch in Appendix B, and for information concerning procedures used in the MPS
- C. Morton Hawkins, for review of the Hypertension Detection and Follow-Up Program sketch in Appendix B
- Charles Hennekens, for review of the Physicians' Health Study sketch in Appendix B
- Fred Heydrick, for critical review of and editorial help on Chapter 21
- Genell L. Knatterud, for review of the University Group Diabetes Program sketch in Appendix B and for information concerning procedures used in the UGDP, DRS, and ETDRS
- William F. Krol, for review of the Persantine Aspirin Reinfarction Study sketch in Appendix B and for information concerning procedures used in PARIS and AMIS
- John M. Lachin, for review of the National Cooperative Gallstone Study sketch and for information concerning procedures used in the NCGS
- John M. Long, for review of the Program on the Surgical Control of Hyperlipidemia sketch in Appendix B
- Maureen Maguire, for help on Chapter 9
- Medical librarians of the Johns Hopkins University, especially Katherine Branch and Karen Higgins, for referencing help throughout the book

- Lisa Mele, for help on Chapter 7
- Larry Moulton, for help on Chapter 9
- Ronald Prineas, for review of the Hypertension Prevention Trial sketch in Appendix B
- Michael Terrin, for help in carrying out the SMOG analysis on consent statements contained in Appendix E
- Robert Weiss, for review of the International Reflux Study in Children sketch in Appendix B

Thanks also go to Sheila Booker, Janet Hiller, Mary Hurt, Joan Jefferys, Jeanette Lautenschlager, Teresa Lee, Mark Van Natta, and Deborah Zeiler for help in producing this manuscript. Sheila Booker deserves a special note of thanks since it is she who did the bulk of the typing, always with great efficiency, accuracy, and good grace, even when working with copy in my worst hand that was almost impossible to decipher, even by me!

A thank you also to colleagues in the School of Hygiene and Public Health for their encouragement and for providing an environment and support structure conducive to a writing effort of this sort. And last I wish to express my gratitude to Jeffrey W. House and Joan Bossert, editors at Oxford University Press in New York. I am especially indebted to Joan Bossert for her exquisite editorial eye and ear and for her professionalism, patience, and persistence in dealing with a stubborn Midwesterner!

Contents

Tables and figures

Part I. Introduction and current status

Chapters in This Part

The seven chapters in this Part cover a number of background issues. The first provides a historical sketch of clinical trials and defines the class of trials considered in this book. Chapter 2 reviews the state of the art of clinical trials, as gleaned from published reports of clinical trials. Chapter 3 defines the stages of activities in a "typical" trial and discusses factors which influence these activities. Chapter 4 provides a definition of single and multicenter trials and discusses a number of issues related to these two classes of trials. Chapter 5 focuses on specialty centers of a multicenter trial with emphasis on coordinating centers. Chapter 6 summarizes available cost data on trials, as provided in the National Institutes of Health Inventory of Clinical Trials, and reviews factors that influence the cost of trials. The last chapter discusses factors that influence the way in which results from trials are viewed and used in everyday medical practice. The University Group Diabetes Program is used as a case study.

1. Introduction

Those who cannot remember the past are condemned to repeat it.

George Santayana

1.1 DEFINITION

A clinical trial is a planned experiment designed to assess the efficacy of a treatment in man by comparing the outcomes in a group of patients treated with the test treatment with those observed in a comparable group of patients receiving a control treatment, where patients in both groups are enrolled, treated, and followed over the same time period. The groups may be established through randomization or some other method of assignment. The outcome measure may be death, a nonfatal clinical event, or a laboratory test. The period of observation may be short or long depending on the outcome measure.

Under this definition, studies involving test and control-treated groups that are treated and followed over different time periods, such as studies involving a historical control group, do not qualify as a clinical trial. Also excluded are comparative studies involving animals other than man, or studies that are carried out in vitro using biological substances from man.

1.2 HISTORY OF CLINICAL TRIALS

The history of clinical trials has been traced by several persons, most notably by Bull (1959) and more recently by Lilienfeld (1982). Table 1-1 provides a summary of some of the historical events in the field of clinical trials.

The concepts involved in clinical trials are ancient. The Book of Daniel, verses 12 through 15, contains an account of a planned experiment with both baseline and follow-up observations.

> *Prove thy servants, I beseech thee, ten days; and let them give us pulse to eat, and water to drink. Then let our countenances be looked upon before thee, and the countenance of the children that eat of the portion of the King's meat: and as thou seest, deal with thy servants. So he consented to them in this matter, and proved them ten days. And at the end of ten days their countenances appeared fairer and fatter in flesh than all the children which did eat the portion of the King's meat (American Bible Society, 1816).*

Avicenna, an Arabian physician and philosopher (980–1037), in his encyclopedic *Canon of Medicine*, set down seven rules to evaluate the effect of drugs on diseases. He suggested that a remedy should be used in its natural state, with uncomplicated disease, and should be observed in two "contrary types of disease." His *Canon* also suggested that the time of action and reproducibility of the treatment effect should be studied (Crombie, 1952).

Many of the early observations affecting choice of treatment were fortuitous and arose from natural consequences rather than planned experiments. The famous observation of the Renaissance surgeon, Ambroise Paré (1510–1590), during the battle to capture the castle of Villaine in 1537, is a case in point (Packard, 1921). Normal treatment procedure for battlefield injuries was to pour boiling oil over the wound. When Paré ran out of oil he found it necessary to resort to an alternative treatment consisting of a digestive made of egg yolks, oil of roses, and turpentine. Paré recognized the superiority of the treatment the next day.

> *I raised myself very early to visit them, when beyond my hope I found those to whom I had applied the digestive medica-*

Table 1-1 Historical events in the development of clinical trials

Date	*Author*	*Event*
1747	Lind	Experiment with untreated control group (Lind, 1753)
1799	Haygarth	Use of sham procedure (Haygarth, 1800)
1800	Waterhouse	U.S.-based smallpox trial (Waterhouse, 1800, 1802)
1863	Gull	Use of placebo treatment (Sutton, 1865)
1923	Fisher	Application of randomization to experimentation (Fisher and MacKenzie, 1923)
1931	—	Special committee on clinical trials created by the Medical Research Council of Great Britain (Medical Research Council, 1931)
1931	Amberson	Random allocation of treatment to groups of patients (Amberson et al., 1931)
1937	—	Start of NIH grant support with creation of the National Cancer Institute (National Institutes of Health, 1981b)
1944	—	Publication of multicenter trial on treatment for common cold (Patulin Clinical Trials Committee, 1944)
1946	—	Promulgation of Nuremberg Code for Human Experimentation (Curran and Shapiro, 1970)
1962	Hill	Publication of book on clinical trials (Hill, 1962)
1962	Kefauver, Harris	Amendments to the Food, Drug and Cosmetic Act of 1938 (United States Congress, 1962)
1966	—	Publication of U.S. Public Health Service regulations leading to creation of Institutional Review Boards for research involving humans (Levine, 1981)
1967	Chalmers	Structure for separating the treatment monitoring and treatment administration process (Coronary Drug Project Research Group, 1973a)
1979	—	Establishment of Society for Clinical Trials (Society for Clinical Trials, Inc., 1980)
1980	—	First issue of *Controlled Clinical Trials*

ment, feeling but little pain, their wounds neither swollen nor inflamed, and having slept through the night. The others to whom I had applied the boiling oil were feverish with much pain and swelling about their wounds. Then I determined never again to burn thus so cruelly the poor wounded by arquebuses (*Packard, 1921*).

An indication that lemon juice was effective in preventing scurvy was the result of a fortuitous decision made by the East India Shipping Company in 1600. Only one of the company's four ships that sailed February 13, 1600, that of General James Lancaster, was supplied with lemon juice. Almost all of the sailors on board Lancaster's vessel remained free of scurvy, while most of the men on board the other three vessels fell victim to the disease. This led shipping company officials to conclude:

And the reason why the General's men stood better in health then [*sic*] *the men of other ships, was this: he brought to sea with him certaine Bottles of the Juice of Limons, which hee gave to each one, as long as it would last, three spoonfuls every morning fasting: not suffering them to eate any thing after it till noone. This juice worketh much the better if the partie keepe, a short Dyet, and wholly refraine salt meate, which salt meate, and long being at the Sea is the only cause of the breeding of this Disease* (*Drummond and Wilbraham, 1940*).

The first planned experiments were done without a formal comparison group. The results of the experiment, contrasted with previous experience, provided the basis for evaluation. The early smallpox experiments are a case in point. A study carried out by Lady Mary Wortley-Montague and Maitland in 1721 involved six inmates from Newgate prison, all assumed to have had no previous exposure to smallpox. The inmates were recruited through a policy, urged by Lady Wortley-Montague, in which King George I commuted the sentence of convicted felons if they agreed to inoculation. The prisoners were inoculated by engrafting smallpox matter from a patient with the natural disease onto both arms and the right leg. The fact that they

remained free of smallpox was taken as evidence in favor of inoculation[1] (Creighton, 1894).

Jenner (1749–1823) described a series of experiments that involved 14 persons, or thereabouts, who had been vaccinated with cowpox (Baron, 1838). He later inoculated three of these people with smallpox and the others with cowpox. He subsequently wrote:

After the many fruitless attempts to give the Small-pox to those who had had the Cow-pox, it did not appear necessary, nor was it convenient to me, to inoculate the whole of those who had been the subjects of these late trials; yet I thought it right to see the effects of variolous matter on some of them, particularly William Summers, the first of these patients who had been infected with matter taken from the cow. He was therefore inoculated with variolous matter from a fresh pustule; but, as in the preceding Cases, the system did not feel the effects of it in the smallest degree (*Jenner, 1798*).

Early experiments with anesthetics (ether and chloroform) in the 1840s by Long, Wells, Morton, and Simpson involved only a few patients and no control group (Duncum, 1947). The ability to render an individual unconscious and then to revive that individual was sufficient to establish the usefulness of anesthetics.

None of the early evaluations of penicillin involved controls. The dramatic recoveries achieved in treating infections, theretofore fatal, were by themselves sufficient to establish the efficacy of the treatment (Keefer et al., 1943).

One of the first experiments designed with a concurrently treated control group involved scurvy victims and was carried out by James Lind in 1747, while at sea on board the *Salisbury*. The study consisted of six different dietary regimens as described by Lind.

On the 20th of May 1747, I took twelve patients in the scurvy, on board the Salisbury *at sea. Their cases were as similar as I could have them. They all in general had putrid gums, the spots and lassitude, with weakness of their knees. They lay together in one place, being a proper apartment for the sick in the fore-hold; and had one diet common to all, viz., watergruel sweetened with sugar in the morning; fresh mutton-broth often times for dinner; at other times puddings, boiled biscuit with sugar, etc.; and for supper, barley and raisins, rice and currants, sago and wine, or the like. Two of these were ordered each a quart of cyder a-day. Two others took twenty-five gutts of* elixir vitriol *three times a-day, upon an empty stomach; using a gargle strongly acidulated with it for their mouths. Two others took two spoonfuls of vinegar three times a-day, upon an empty stomach; having their gruels and their other food well acidulated with it, as also the gargle for their mouth. Two of the worst patients, with the tendons in the ham rigid, (a symptom none of the rest had), were put under a course of sea-water. Of this they drank half a pint every day, and sometimes more or less as it operated, by way of gentle physic. Two others had each two oranges and one lemon given them every day. These they eat with greediness, at different times, upon an empty stomach. They continued but six days under this course, having consumed the quantity that could be spared. The two remaining patients, took the bigness of a nutmeg three times a-day, of an electuary recommended by an hospital surgeon, made of garlic, mustard-seed, rad raphan, balsam of Peru, and gum myrrh; using for common drink, barley-water well acidulated with tamarinds; by a decoction of which, with the addition of* cremor tartar, *they were gently purged three or four times during the course.*

Those receiving a daily ration of oranges and lemons fared best.

The consequence was, that the most sudden and visible good effects were perceived from the use of the oranges and lemons; one of those who had taken them, being at the end of six days fit for duty (*Lind, 1753*).

Still, in spite of these findings, Lind and others clung to the notion that the best treatment involved placing patients stricken with scurvy in "pure dry air." The reluctance to accept oranges and lemons as treatment for the disease had to do, in part, with the relative expense of acquiring such fruits as opposed to the "dry air" treatment. It was 1795 before the British Navy supplied lemon juice for its ships at sea (Drummond and Wilbraham, 1940).

1. The results were not as convincing as first perceived. One of the six inmates was subsequently found to have had smallpox before inoculation and a second may have had the disease in childhood (Creighton, 1894).

The importance of a control treatment as a means of identifying placebo effects was recognized by Haygarth (1740–1827) in his 1799 study of Perkin's Tractors—metallic rods used to stroke the body of an ailing person (Haygarth, 1800). The rods were widely used at the time for a variety of conditions, including crippling rheumatism, pain in the joints, wounds, gout, pleurisy, and inflammatory tumors, as well as for "sedating violent cases of insanity." Haygarth used imitation tractors made of wood on five patients affected with chronic rheumatism.

> *Let their [the Tractors'] merit be impartially investigated, in order to support their fame, if it be well founded, or to correct the public opinion, if merely formed upon delusion. Such a trial may be accomplished in the most satisfactory manner, and ought to be performed without any prejudice. Prepare a pair of false, exactly to resemble the true Tractors. Let the secret be kept inviolable, not only from the patient, but every other person. Let the efficacy of both be impartially tried; beginning always with the false Tractors. The cases should be accurately stated, and the reports of the effects produced by the true and false Tractors be fully given, in the words of the patients. . . .*
>
> *On the 7th of January, 1799, the wooden Tractors were employed. All the five patients, except one, assured us that their pain was relieved. . . .*

The following day Haygarth used the metallic tractors on the same patients. He observed:

> *All the patients were in some measure, but not more relieved by the second application, except one, who received no benefit from the former operation, and who was not a proper subject for the experiment, having no existing pain, but only stiffness of her ankle (Haygarth, 1800).*

Sir William Gull (1816–1890), in collaboration with Henry Sutton, demonstrated the importance of placebo treatment in assessing the natural variability of the course of disease and the possibility of spontaneous cure. They gave mint water to 44 rheumatic fever patients and, after close observation, concluded:

> *The cases show that too much importance has been attached to the use of medicines, especially those acute cases where the tendency to a natural cure is the greatest (Sutton, 1865).*

Most of the early experiments involved arbitrary, nonsystematic schemes for assigning patients to treatment, such as that described by Lind. More systematic approaches were needed for trials in which patients were enrolled in a sequential fashion. Johannes Fibiger, in an evaluation of a therapeutic serum for the treatment of diphtheria patients, used a scheme in which "serum was injected into all those admitted on every other day" (Fibiger, 1898). Park and co-workers, in 1928, described a scheme involving use of an experimental treatment for lobar pneumonia on every other patient.

> *Patients were therefore taken alternatively for antibody treatment or control depending only on the order of their admission to the service. It was believed that with a sufficiently large series the distribution of cases by type would be equalized between the treated and the untreated group (Park et al., 1928).*

The concept of randomization as a device for treatment assignment was introduced by Fisher while he was involved in agricultural experimentation (Box, 1980; Fisher and MacKenzie, 1923; Fisher, 1926, 1973). Amberson and his co-workers, in a study of sanocrysin in the treatment of pulmonary tuberculosis, were among the first to use the concept for treatment assignment in an actual clinical trial.

> *The 24 patients were then divided into two approximately comparable groups of 12 each. The cases were individually matched, one with another, in making this division. . . . Then, by a flip of the coin, one group became identified as group I (sanocrysin-treated) and the other as group II (control). The members of the separate groups were known only to the nurse in charge of the ward and to two of us. The patients themselves were not aware of any distinctions in the treatment administered (Amberson et al., 1931).*

It was several years later before the process of randomization was used for assigning individual patients to treatment. Diehl and co-workers (1938) described a method of randomly assigning University of Minnesota student volunteers to treatment in a double-masked, placebo-controlled trial involving treatment of the common cold.

Great Britain, under the influence of men such as Sir Austin Bradford Hill, has been a leading force in the development of modern-day clinical trials. His book *Statistical Methods in Clinical and Preventive Medicine* (1962) represents an important milestone in the field of clinical trials.

The Medical Research Council of the United Kingdom recognized the need for clinical trials at least as early as 1930. An announcement in a 1931 issue of *Lancet* stated:

> *The Medical Research Council announce that they have appointed a Therapeutic Trials Committee, as follows, to advise and assist them in arranging for properly controlled clinical tests of new products that seem likely, on experimental grounds, to have value in the treatment of disease. . . . The Therapeutic Trials Committee will be prepared to consider applications by commercial firms for the examination of new products, submitted with the available experimental evidence of their value, and appropriate clinical trials will be arranged in suitable cases (Medical Research Council, 1931).*

The concept of multiple investigators from different sites, all following a common study protocol in the conduct of a clinical trial, did not emerge until the late 1930s and early 1940s. One of the first applications of this approach appeared in a 1944 publication of a trial to evaluate patulin for treatment of the common cold (Patulin Clinical Trials Committee, 1944).

A multicenter trial involving the use of streptomycin in patients with pulmonary tuberculosis was published in 1948 (Medical Research Council, 1948). One of the first multicenter trials in the United States involved assessment of the same drug (Mount and Ferebee 1952, 1953a, 1953b). The study was initiated about the same time as the British study but did not produce any published results until 1952—four years after the British publication.

The Veterans Administration (VA), in conjunction with the United States Armed Services, carried out a series of multicenter trials between 1945 and 1960 in an attempt to establish the efficacy of various chemotherapeutic agents in the treatment of tuberculosis (Tucker, 1960). The VA provided support for various other multicenter trials in the 1960s under a relatively informal funding structure. A more formal structure was created in 1972.

The United States poliomyelitis vaccine trials, started in the autumn of 1953, sponsored by the National Foundation for Infantile Paralysis and done in collaboration with the Public Health Service and state health departments, were multicenter (Francis et al., 1955). They are noteworthy because of their size. They involved tens of thousands of volunteers.

The creation of the National Cancer Institute in 1937 signaled the start of federally sponsored medical research in the United States and the creation of what ultimately has come to constitute the National Institutes of Health (National Institutes of Health, 1981b). The Institutes of this agency support by far the largest number of trials among all United States governmental agencies. The largest and most complex multicenter trials have been carried out by the National Heart, Lung, and Blood Institute (NHLBI). Some, such as the Multiple Risk Factor Intervention Trial (Multiple Risk Factor Intervention Trial Research Group, 1977) and the Hypertension Detection and Follow-Up Program (Hypertension Detection and Follow-Up Program Cooperative Group, 1979a), have involved thousands of patients and years of follow-up.

One of the first multicenter trials sponsored by the National Heart Institute (now the National Heart, Lung, and Blood Institute) was a trial involving the use of ACTH, cortisone, and aspirin as a treatment for rheumatic heart disease. The trial was initiated in 1951 and was carried out in conjunction with the Medical Research Council of Great Britain, the American Heart Association, and the Canadian Arthritis and Rheumatism Society (Rheumatic Fever Working Party, 1960).

Multicenter trials, focusing on the treatment of chronic noninfectious diseases, began to appear in the 1960s. One of the first examples in this category was the University Group Diabetes Program, started in 1960 and completed in 1974 (University Group Diabetes Program Research Group, 1970e, 1978).

The advent of multicenter clinical trials as a treatment evaluation tool has required collaboration among various disciplines. In addition to medical and biostatistical expertise, a typical large-scale multicenter trial requires close participation with various other specialists. This multidisciplinary approach has served to stimulate communication across disciplines, as evidenced by formation of the Society for Clinical Trials in 1979 and publication of *Controlled Clinical Trials* starting in 1980.

A major stimulus for the execution of clinical trials in the United States arose from language included in the 1962 Kefauver-Harris amendments to the United States Food, Drug and Cosmetic Act of 1938. The Act set forth a series of legal requirements which had to be satisfied before a drug could be approved by the Food and Drug Administration—FDA (Colsky, 1963; Food and Drug Administration, 1963; Kelsey, 1963; United States Congress, 1962). A unique feature of the amendment was language spelling out the nature of scientific evidence required for a drug to be approved for human use—a specification heavily dependent on what are referred to in the act as "adequate and well-controlled investigations."

> *The term "substantial evidence" means evidence consisting of adequate and well-controlled investigations, including clinical investigations, by experts qualified by scientific training and experience to evaluate the effectiveness of the drug involved, on the basis of which it could fairly and responsibly be concluded by such experts that the drug will have the effect it purports or is represented to have under the conditions of its use prescribed, recommended, or suggested in the labeling or proposed labeling thereof* (*United States Congress, 1962*).

Regulations published in the *Federal Register* (Food and Drug Administration, 1969a, 1969b, 1970a, 1970b) have set forth general design and execution standards for trials carried out as part of a FDA Investigational New Drug Application (INDA) and New Drug Application (NDA) processes. They were taken in large measure from testimony given by William Beaver in a court case involving the Pharmaceutical Manufacturers Association versus Robert H. Finch, Secretary of Health, Education and Welfare, and Herbert L. Ley, Commissioner of Food and Drugs (Crout, 1982; United States District Court, 1969, 1970).

The Medical Device Amendments of 1976 have extended some of the testing requirements established for drugs to medical devices as well (United States Congress, 1976). Certain devices cannot be marketed without supporting evidence of safety and efficacy as obtained through controlled trials.

The importance of safe and effective treatments for major diseases has led Congress to earmark money for targeted areas of research. The Coronary Drug Project (CDP) is an early example of a trial funded via this route (Coronary Drug Project Research Group, 1973a). The emphasis on focused research has led to increased use of research contracts in place of grants by the NIH as funding vehicles for many of the large-scale multicenter trials (see Chapters 5 and 21).

The long-term multicenter trial has created a new class of organizational and analysis problems. A special task force convened by the National Heart Institute in 1967 outlined organization guidelines that have been used for many of the large-scale trials since then (Greenberg, 1967).[2] The analytic problems created by the need for periodic data analyses as the trial proceeds have led to the development of organizational structures that provide for a separation of the patient care and treatment evaluation functions. The structures, described in Chapter 23, emerged from concerns regarding the possibility of bias if study physicians are permitted access to study data during the course of the trial (Meinert, 1981). Chalmers was an early proponent of this separation of functions in the organization of the CDP.[3]

Cornfield played a major role in developing a philosophy that dealt with the problems of ongoing analyses in long-term clinical trials (Greenhouse and Halperin, 1980; Seigel, 1982). His work on Bayesian analysis and on the use of the likelihood principle as an analytic tool served to de-emphasize the role of significance testing in data evaluation (Cornfield, 1969).

1.3 TERMINOLOGY CONVENTIONS

The language of clinical trials is confusing. Language conventions have not been established for characterizing the key design, organizational, and operational elements of trials (Meinert, 1980a). Appendix A provides a glossary of terms, abbreviations, and acronyms used in this book.

The term *patient* (see Glossary for the derivation) will be used throughout to denote an individual enrolled in a trial. It will be used even though it may not always be appropriate, for example, as in trials that involve people without clinical disease. The term *test treatment* will denote the treatment to be evaluated in the trial. The term *control treatment* will denote the treatment used for comparison with the test treatment. For convenience, study designs will be discussed as if they involve a single test and control treatment, although certain trials may involve several test treatments. The term *study treatments* will denote the entire set of test and control treatments used in a trial.

The term *trial* is from the Anglo-French word *trier*, meaning to choose, sort, select, or try (Klein, 1971). Thomas Bayes (1702-1761), an English mathematician, made frequent use of the term in a nonmedical experimental sense in an essay on probability involving repeated drops of a billiard ball onto a surface to observe the position of its fall (Bayes, 1763). The use of the term in a medical context is not easy to trace. However, even a cursory search indicates it has been in use for some time. It appears in the writings of both Haygarth and Jenner around 1800. Its use today covers a wide variety of designs ranging from uncontrolled observations involving the first use of a treatment in man to a formal experiment, complete with a control treatment and randomization. The use of the term without modifiers implies nothing about the observational unit. It may be man or some other animal species—always man in this book.

Trial is frequently modified by the term *clinical* and/or one or more design terms (e.g., *randomized, placebo, controlled,* or *double-blind*). Table 1-2 provides an indication of modifier usage as seen in 1980 nonreview, publications in English appearing in the MEDLINE[4] data file. The results presented are for articles appearing under the check tag *human*—a designation applied by indexers at the National Library of Medicine (NLM) to identify studies involving humans.[5] Tabulations presented in the first column of the table are based on a search of all the titles indexed under the medical subject heading (MeSH) *clinical trials* (1,949). Of the 502 articles containing the term *trials*, 40% also contained the term *clincial*. The term *trial* appeared without any of the modifiers listed in Table 1-2 in 20% of the titles (99 out of 502). It is worth noting that nearly three-fourths of the 1,949 articles screened did not contain the term *trial*. Other more nondescript terms such as *study* were used instead (see Chapter 2 and Coordinating Center Models Project Research Group, 1979e). Unfortunately, this pattern of use creates problems when an attempt is made to identify trials via title searching routines.

The results in the last column in Table 1-2 concern the use of the term *trial(s)* in articles appearing under MeSH headings other than *clinical trials*. A number of these may very well involve studies that are nonexperimental. Theoretically, this should be true for all articles not classified under the MeSH *clinical trials*. However, some of the articles identified appear to be germane to the field, as suggested by use of modifiers such as *clinical, controlled, double-blind, random,* or *randomized.*

Table 1-2 Frequency of selected terms in titles published in 1980*

Term used	Titles under: MeSH of clinical trials		Titles under: Other MeSH	
Titles containing the term *trial(s)*	502	(100)	191	(100)
Titles containing the term *trial(s)* plus:				
Clinical	201	(40)	41	(22)
Controlled	131	(26)	23	(12)
Double-blind	79	(16)	16	(8)
Random(ized)	74	(15)	13	(7)
Comparative	22	(4)	9	(5)
Field	15	(3)	4	(2)
Titles containing the term *trial(s)* and none of the above terms	99	(20)	103	(54)

*MEDLINE search, as of June 1982. Run restricted to nonreview articles in English appearing under the check tag *human*.

2. This report, according to William Zukel of the NHLBI (personal communication, 1982), drew heavily on organizational experience gained from earlier multicenter studies, most notably those done by the Committee on Lipoproteins (1956) and by the Rheumatic Fever Working Party (1960).

3. A written communication from Thomas Chalmers to the Chairman of the CDP Policy Board, Robert Wilkens, in 1967, led to the separation of these functions in the CDP.

4. Medical Literature Analysis Retrieval System On Line, a computer database of literature citations produced by the National Library of Medicine (Williams et al., 1979).

5. Most of the articles under the heading *clinical trials* appear under this tag. However, beginning in January 1981, the heading includes veterinary studies and hence contains studies where only the check tag *animal* is used.

1.4 FOCUS

This book will focus on the class of trials that involve:

- Man
- A fixed, nonsequential sample size design
- Random allocation of individual patients to treatment, as opposed to some larger randomization unit such as family, hospital ward, community, etc.
- An uncrossed treatment design (i.e., where the treatment design requires patients to receive either the test or control treatment, but not both)
- Concurrent enrollment, treatment, and follow-up of patients in the test and control treatment groups
- A clinical event, such as death or some other nonfatal event (e.g., a myocardial infarction, recurrence of cancer, loss of vision, etc.), as the primary outcome measure for evaluating the test treatment

The fixed sample size design is by far the most commonly used design for the class of trials considered. Sequential designs (see Chapter 9 for further discussion) are not practical for comparing treatments in trials requiring long periods of follow-up for outcome assessment.

Emphasis will be on trials that require multiple clinics in order to enroll the required number of patients (see Appendix B for examples). The researcher who can cope with the challenges presented by such trials is in a good position to deal with less complicated trials carried out in a single clinic.

Many of the principles discussed herein have applicability beyond the setting outlined. This is true for several of the chapters, particularly those concerned with data collection (Chapter 12) and with organization and management practices (Chapters 22 and 23).

2. Clinical trials: A state-of-the-art assessment

One's knowledge of Science begins when he can measure what he is speaking about and express it in numbers.

Lord Nelson

2.1 EXISTING INVENTORIES

Various groups have assumed responsibility for developing and maintaining inventories of ongoing clinical trials. Some are organized according to disease; others relate to trials sponsored by a specific agency. An early example of the first type of inventory originated from the National Institute of Mental Health with the creation of the Biometric Laboratory Information Processing System (BLIPS) in the mid-1960s. The inventory was created to provide information for ongoing trials of psychopharmacological agents in the United States and elsewhere (Levine et al., 1974). The National Cancer Institute (1983), via the International Cancer Research Data Bank, maintains a worldwide file of ongoing phase II and phase III cancer trials. The Veterans Administration (VA) maintains a list of trials carried out under its collaborative studies program (list available from the VA Central Office, 810 Vermont Avenue N.W., Washington, D.C.).

The Division of Research Grants of the National Institutes of Health (NIH) has maintained an inventory of NIH-sponsored trials for several years (National Institutes of Health, 1975, 1980). Responsible officials of institutes of the NIH involved in extramural or intramural research are asked to complete inventory sheets for all ongoing studies that they consider to satisfy the definition of a clinical trial, as specified in the inventory. The definition used is:

> *A scientific research activity undertaken to define prospectively the effect and value of prophylactic/diagnostic/therapeutic agents, devices, regimens, procedures, etc., applied to human subjects. It is essential that the study be prospective, and that intervention of some sort occur. The choice of number of cases or patients will depend on the hypothesis being tested, but must be sufficient to permit a definite result to be anticipated. Phase I, feasibility, or pilot studies are excluded.*

This definition allows inclusion of trials with only one treatment group. One can only surmise that evaluation of the treatment is made against some hypothetical standard control treatment or through use of historical controls in such cases (see Chapter 1 and Glossary for definition of *clinical trial* as used in this book). The broad nature of the definition and the lack of surveillance by the Division of Research Grants in monitoring for differences in how the definition is applied allows for considerable variability in the reporting behavior of institutes contributing to the inventory. It is likely that some of the variation among institutes, within and across years, evident in tables in this chapter and in Chapters 5 and 6, is due to differences in reporting practices. Unfortunately, the inventory is not designed to provide data on the nature of the differences.

The number of trials reported for the 5-year period for which inventory data are available ranged from a low of 746 in 1977 to a high of 986 in 1979 (Table 2-1). The "typical" NIH trial, as reflected in the 1979 NIH Inventory,[1] involved between 30 and 300 patients (median sample size: 100) apportioned among the different treatment groups (Table 2-2). Most of the trials were classified as therapeutic (81%), as compared to prophylactic (13%) and diagnostic (6%) (see Glossary for definitions). The majority (65%) were funded for a period of 3 years or longer.

Trials sponsored by the individual institutes vary in number and size (Table 2-3). The National Cancer Institute (NCI) sponsored by far the most trials (654 out of the 985 listed for 66% of all NIH trials). The National Heart, Lung, and Blood Institute (NHLBI) sponsored the largest trials (median sample size: 850). This variation in size is due, in part, to differences in the nature of the health problems addressed. The NCI plays a major role in developing and testing chemotherapeutic agents. Hence, many of their trials are of the phase I or II variety (see Glossary), involving relatively small numbers of patients. The NHLBI has concentrated on assessing the usefulness of various drugs and procedures in the primary or secondary prevention of heart disease. Their trials, of necessity, have had to involve large numbers of patients and long periods of follow-up because of low underlying event rates for the outcomes of interest.

1. There have been no inventories since 1979, but one is planned or 1984 or 1985.

Table 2-1 Number of trials, median sample size, and percent randomized by fiscal year, as reported in NIH Inventories of Clinical Trials

Fiscal year	*Total number of trials*	*Median sample size*	*Percent randomized*
1975	755	127	62
1976	926	114	60
1977	746	125	62
1978	845	103	60
1979	986	100	60

Table 2-2 Design features of trials reported in the 1979 NIH Inventory of Clinical Trials

Design features	*Number of trials*	*Percent*
Number of treatment groups per trial		
1	258	26
2	438	44
≥3	290	29
Median number of treatment groups/trial: 1.48		
Sample size		
Median number of patients/trial	100	
Range (20th to 80th percentile)	30 to 300	
Number of patients/trial/treatment group*	68	
Range (20th to 80th percentile)	20 to 203	
Method of treatment allocation		
Random	589	60
Nonrandom	391	40
Method not reported	6	0
Type of trial		
Therapeutic	801	81
Prophylactic	126	13
Diagnostic	58	6
Anticipated length of trial		
≤1 year	19	2
1 year to ≤2 years	101	10
2 years to ≤3 years	223	23
>3 years	642	65
Total number of trials listed	986	100

*Calculated by dividing median number of patients per trial by the median number of treatment groups per trial.

Table 2-3 Number of trials, median sample size, and percent randomized, as reported in the 1979 NIH Inventory of Clinical Trials

Institute	*Number of trials*	*Median sample size*	*Percent randomized*
National Cancer Institute (NCI)	654	100	59
National Eye Institute (NEI)	26	200	85
National Heart, Lung, and Blood Institute (NHLBI)	20	850	100
National Institute of Allergy and Infectious Diseases (NIAID)	120	100	53
National Institute of Arthritis, Metabolism, and Digestive Diseases (NIAMDD)	67	70	60
National Institute of Child Health and Human Development (NICHD)	32	100	62
National Institute of Dental Research (NIDR)	26	663	65
National Institute of Neurological and Communicative Disorders and Stroke (NINCDS)	40	30	55
Total	985*	100	60

*One trial sponsored by the National Institute of General Medical Services not included.

2.2 TRIALS AS SEEN THROUGH THE PUBLISHED LITERATURE

An indication of the nature of completed trials can be obtained from a review of the published literature, as identified through *Index Medicus* or MEDLINE—the computerized version of *Index Medicus* (Beatty, 1979; Charen, 1977; Kenton and Scott, 1978; McCarn, 1980; Williams et al., 1979). The introduction in 1980 of a subject heading for *clinical trials* has made it possible to retrieve articles under this heading.[2] The definition used by indexers at the National Library of Medicine—the agency responsible for entries into *Index Medicus* and MEDLINE—is:

> *Pre-planned usually controlled studies of the safety, efficacy, or optimum dosage schedule (if appropriate) of one or more diagnostic, therapeutic, or prophylactic drugs or technics in humans selected according to pre-determined criteria of eligibility and observed for pre-defined evidence of favorable and unfavorable effects (National Library of Medicine, 1980).*

This definition, as with the one used by NIH, is designed to permit inclusion of studies with a wider number of design features, including some without a comparison group.

The heading included 2,409 citations bearing a 1980 publication date, as of an October 1981 MEDLINE search.[3] This number represents less than 1% of the total 1980 MEDLINE citations (Table 2-4). The 1,796 titles remaining after exclusion of review articles and foreign-language papers were ordered by date of entry into the MEDLINE file (approximately chronological by date of publication) and then sampled using a random start and a 1 in 10 sampling fraction. A total of 67 (37%) of the 180 papers selected were

2. Before 1980, trials were classified under the general heading *clinical research*.

3. This run included most of the 1980 publications. Evidence from previous years indicates that 95% of all entries for a given calendar year are indexed and entered into the system by October of the following year.

Table 2-4 1980 publications cited in MEDLINE as of October 1981

249,150	Total number of 1980 entries in MEDLINE
2,409	Number of 1980 titles under heading *clinical trials*
2,317	Number of 1980 titles remaining after exclusion of review articles
1,796	Number of 1980 titles remaining after the exclusion of review and foreign-language publications

eliminated for reasons indicated in Table 2-5. The tabulations given in Tables 2-6 through 2-9 are based on the 113 remaining papers. Appendix C contains a list of all 180 papers (see also Meinert et al., 1984).

It would have been necessary to subscribe to no less than 82 different journals in order to have access to the 113 articles reviewed. Moreover, no combination of 4 or 5 journals accounted for a majority of the articles. Only 17 of the 82 journals contained 2 or more of the papers selected for review (Table 2-6). The 8 most frequently cited journals accounted for a little more than a quarter (27%) of the 113 articles (Table 2-7).

Each paper in the sample was classified as to major subject area (Table 2-8). General design characteristics of the trials represented in the sample are summarized in Table 2-9. The typical trial, as seen through published literature, is carried out in a single clinic and involves about 25 patients per treatment group followed over a relatively short time—usually less than 3 months.

Table 2-5 Literature selection process for papers appearing under heading *clinical trials*

Total number of English, nonreview, 1980 publications	1,796	
Number of papers selected in sample	180	
Number of papers excluded after initial review	67	
No comparison group		15
Editorial or letter		17
Review or methodological paper		24
Other reasons*		11
Number reviewed	113	

*Includes 1 paper that could not be located, 8 position or philosophical papers, and 2 others not classified as clinical trials under the definition used in this book (see Chapter 1 and Glossary).

Table 2-6 Number of journals represented in sample of 113 papers

Number of journals represented in sample of 113 papers	[illegible]
Number of journals with:	
1 of the 113 papers	[illegible]
2 of the 113 papers	[illegible]
3 or more of the 113 papers (see Table 2-7)	[illegible]

Source: Reference citation 321. Reprinted with permission of Elsevier Science Publishing Co., Inc., New York.

Over 70% of the trials were classified as therapeutic (see Glossary for definition). The overwhelming proportion of trials involved drug treatments. Only 10 of the 113 studies involved some other form of treatment. Of the 10, 5 were surgical trials, 2 involved behavior modification, 2 involved a radiologic procedure, and 1 involved testing a medical device. Approximately a third (31%) of the trials used crossover designs (see Glossary).

The median length of follow-up was slightly over 2 months. There were only 12 trials that provided for a year or more of follow-up. Over two-thirds of the trials were reported to be double-masked; 80% of the reports indicated use of some random method for treatment assignment. Treatment assignment was classified in the nonrandom or unstated category if the paper contained an explicit statement indicating use of a nonrandom method or if there was no way to determine how assignments were made.

Fifteen of the trials (13%) were classified as multicenter. The remainder were classified as single-center (89 studies) or could not be classified because of lack of information in the papers (9 studies). Slightly over half of the papers (53%) indicated a source of funding. Acknowledgment of a contribution of a supply item, such as drugs, was ignored in the classification, unless there was evidence that money was also provided.

Table 2-7 Journal of publication for 113 papers reviewed

Journal	*Number of papers*
Br Med J	[illegible]
Lancet	[illegible]
J Clin Pharmacol	[illegible]
Br J Clin Pharmacol	[illegible]
Br J Dis Chest	[illegible]
Cancer	[illegible]
J Int Med Res	[illegible]
S Afr Med J	[illegible]
All other journals (74)	[illegible]
Total number of papers in sample	[illegible]

Source: Reference citation 321. Reprinted with permission of Elsevier Science Publishing Co., Inc., New York.

Table 2-8 Subject matter of 113 papers reviewed

Subject	*Number of papers*
Cardiovascular	14
Gastrointestinal	14
Psycho-neurological	13
Cancer	10
Urinary system	8
Bone and joint	7
Dermatology	6
Dental	5
Respiratory	5
Allergy	4
Gynecologic	4
Ophthalmologic	3
Pain relief	3
Infectious disease	3
Other*	14
Total number of papers in sample	113

Source: Reference citation 321. Reprinted with permission of Elsevier Science Publishing Co., Inc., New York.

*Anesthesiology; ear, nose, and throat; diabetes; contraception; genetics, diagnostic; trauma; drugs; and weight control.

Most trials presented results for a number of outcome measures (see Glossary). Many of the papers presented results for several different outcomes. It was impossible in nearly all those cases to identify the measure considered to be primary (see Glossary for definition). Most measures were of a nonclinical nature (e.g., usually laboratory or physiological measures). Only 3 trials used mortality as an outcome measure.

2.3 SMALL SAMPLE SIZE: A COMMON DESIGN FLAW

Only 2 of the 113 trials showed any evidence of a sample size calculation and they involved sequential designs. Of the others, 3 mentioned the statistical power (see Glossary) associated with the trial. The virtual disregard of power considerations is consistent with other literature reviews. None of the 83 gastrointestinal trials reviewed by Chalmers and co-workers (1978) included any discussion of power. Only 2 of the 93 papers from breast cancer trials reviewed by Mosteller and co-workers (1980) contained a discussion of power. Power considerations are especially important in trials where investigators conclude in favor of the null hypothesis (Freiman and co-workers, 1978).

2.4 FUTURE NEEDS

The annals of medicine are filled with accounts of potions, drugs, devices, and the like, that have been heralded as great advances only to be shown as useless or even harmful later on. Bloodletting (venesection) has been used therapeutically as well as prophylactically from prehistoric times to the 1950s (Bryan, 1964; Holman, 1955; King, 1961). The death of George Washington was presumably associated with bloodletting (Donaldson and Donaldson, 1980; Knox, 1933). It fell from favor as a treatment for hypertension, not so much because of concerns regarding efficacy of the treatment, but rather, because of the advent of other modes of therapy. Holman, as late as 1955, after a review of medical texts in use at that time, wrote:

> *Bloodletting is still mentioned for control of arterial hypertension. . . . Hypertensive patients not in circulatory failure have often been observed to get symptomatic relief from venesection for varying periods of time. . . . If the early promise of Rauwolfia and similar recently introduced antihypertensive agents is fulfilled, this indication for venesection is apt to be supplanted also.*

Perkin's tractors, introduced in 1795 and mentioned in Chapter 1, continued to be used long after Haygarth's study in 1800 showed them to be of no value (Elliott, 1913; Haygarth, 1800). Nathan Smith, the founder of the Yale Medical School, not only gave testimony to their efficacy but was reported to have sold them (Haggard, 1932).

Changes in treatment philosophy are slow to occur, especially if the new philosophy must replace an established one. Max Planck (1858–1947), a physicist, noted that:

> *A new scientific truth does not triumph by convincing its opponents and making them see the light, but rather because its opponents eventually die, and a new generation grows up that is familiar with it (Strauss, 1968).*

The promotion and use of ineffective treatments is not simply a mistake of the past, as is

Table 2-9 Design characteristics of sample of 113 trials appearing in 1980 published literature

Design characteristic	*Number of trials*	*Percent*
Number of treatment groups		
2	70	62
3	24	21
≥4	19	17
Type of trial		
Therapeutic	81	72
Prophylactic	21	19
Diagnostic	2	2
Uncertain	9	8
Treatment design		
Drug trials	103	91
Uncrossed treatment	68	66
Crossed treatment	32	31
Treatment structure unclear	3	3
Other trials	10	9
Sample size		
≤20	19	17
21-49	36	32
50-99	21	19
100-299	20	18
≥300	15	13
Unstated	2	2
Median number: 52.5 (range 4 to 3,427)		
Median number per treatment group: 26.2 (range 2 to 1,714)		
Length of follow-up		
≤1 week	22	19
>1 week but ≤1 month	20	18
>1 month but ≤3 months	26	23
>3 months but ≤1 year	19	17
>1 year	12	11
Not stated	14	12
Median: 2.1 months (range <1 day to >2 years)		
Method of treatment assignment		
Random	90	80
Nonrandom or not stated	23	20
Level of treatment masking		
Double-masked	76	67
Single-masked	4	4
Unmasked	17	15
Not stated	16	14
Number of centers		
Single center*	98	87
Multicenter	15	13
Type of funding		
Public	24	21
Private	22	19
Public and private	14	12
Not stated	53	47

Source: Reference citation 321. Reprinted with permission of Elsevier Science Publishing Co., Inc., New York.

*This category includes 9 trials with inadequate information to make a classification.

evident from work of the Drug Efficacy Study Implementation, DESI (Food and Drug Administration, 1972b). Of the 3,185 prescription drugs reviewed by the FDA as of June 1982, 31% were classified as ineffective.[4]

The adoption of treatments as established forms of therapy without adequate testing applies to nondrug forms of therapy as well. Coronary artery bypass surgery was introduced in 1964 (DeBakey and Lawrie, 1978; Garrett et al., 1973). Since that time it has become one of the most common forms of surgery performed. Only recently have trials been mounted to evaluate the efficacy of the operation (Braunwald, 1977; Coronary Artery Surgery Research Group, 1981, 1983; European Coronary Surgery Study Group, 1982b; Murphy et al., 1977).

Coronary care units, regarded as standard treatment for patients with myocardial infarction since their introduction in 1962, have never been adequately evaluated (Day, 1965; Gordis et al., 1977). The few controlled trials that have been done raise doubts concerning widespread use of such units (Christiansen et al., 1971; Hill et al., 1977, 1978; Mather et al., 1971, 1976).

The development of electronic fetal monitoring (EFM) devices in the late 1960s has led to their widespread use in delivery rooms. Their use has been accompanied by a marked rise in cesarean section rates, without any apparent improvement in neonatal outcome (Haupt, 1982; Ott, 1981). All of the randomized trials reported to date have failed to show any benefit for the EFM devices tested (Haverkamp et al., 1976, 1979; Kelso et al., 1978; Renou et al., 1976). However, those results have not had any apparent effect on the use of the devices.

Demands from the public for access to new "miracle" drugs can also influence health care practices. Public clamor for Laetrile has led state legislators in 26 states to enact laws making the drug available to the public,[5] in spite of a skeptical medical profession and trials failing to indicate any merit for the treatment (Bross, 1982; Moertel et al., 1982; Relman, 1982). Lobbying by lay groups for a relaxation of proscriptions against the use of dimethyl sulfoxide (DMSO) has led to availability of the compound in 9 states[5] even though there are serious doubts regarding its usefulness (National Research Council, 1973).

The need for clinical trials is not limited to the medical profession. A case in point is the widespread and often indiscriminate use of diethylcarbamazine to protect dogs against heart worms. The risks associated with the chronic use of such medications, year in and year out for the life of a dog, may be greater than the risk from the heart worm itself, especially if the animal lives in a low infestation area and spends most of its time indoors.

The clinical trial has been termed the "indispensable ordeal" by Fredrickson (1968). Indeed it is, if we are to eliminate the uncertainty that stems from lack of data needed to evaluate the merit of many of our current treatment practices.

4. Personal communication with staff of the Office of the Division of Federal and State Relations, Food and Drug Administration, 1982.

5. Personal communication with staff of the Office of the Division of Federal and State Relations, Food and Drug Administration, 1982.

3. The activities of a clinical trial

Field trials are indispensable. They will continue to be an ordeal. They lack glamor, they strain our resources and patience, and they protract the moment of truth to excrutiating limits. Still, they are among the most challenging tests of our skills. I have no doubt that when the problem is well chosen, the study is appropriately designed, and that when all the populations concerned are made aware of the route and the goal, the reward can be commensurate with the effort. If, in major medical dilemmas, the alternative is to pay the cost of perpetual uncertainty, have we really any choice?

Donald Fredrickson (1968)

3.1 STAGES OF A CLINICAL TRIAL

A clinical trial progresses through a series of stages from beginning to end. The stages discussed in this book are outlined in Table 3-1, along with the event that is used to designate the end of one stage and the start of the next. The dates listed in the last column of the table are from the CDP (Coronary Drug Project Research Group, 1973a, 1976).

Appendix D provides a listing of activities by stage. The list is an adaptation of one developed as part of the Coordinating Center Models Project—CCMP (Coordinating Center Models Project Research Group, 1979d). It should be used only as a rough guide to activities in specific trials. It has been constructed assuming no overlap of activities from one stage to the next. In actual fact, as noted in Section 3.3.3, the overlap can be quite extensive.

3.2 DIVISION OF RESPONSIBILITIES

Any trial involving two or more investigators, whether done at a single center or multiple centers, must provide for a division of responsibilities. Some responsibilities, such as those related to patient care or to data analysis, require specialized skills associated with a particular discipline and may automatically be assumed by persons trained in that discipline. However, many of the required functions are not uniquely associated with a specific discipline and can be performed by any one of several individuals or groups in the trial. This fact was evident in the review of the data coordinating centers carried out as part of the CCMP. All centers had the responsibility for data intake and analysis, but they showed wide variation in the number of other general support functions performed. In some trials, the center had responsibility for virtually all support functions, whereas in others responsibilities were shared with or assumed by individuals or groups outside the center (McDill, 1979).

It is useful to list required activities and the individual or group expected to perform them. This should be done early in the trial to avoid confusion as to who is doing what. These specifications are especially important in trials with multiple resource centers that have overlapping responsibilities (McDill, 1979). The specifications, once developed, should be reviewed and revised at intervals over the course of the trial to cover new responsibilities and to realign old ones.

Table 3-1 Stages of a clinical trial

Stage	Event marking end of stage	Illustration using CDP*
I. Initial design	Initiation of funding	March 1965
II. Protocol development	Initiation of patient recruitment	March 1966
III. Patient recruitment	Completion of patient recruitment	October 1969
IV. Treatment and follow-up	Initiation of patient close-out	May 1974
V. Patient close-out	Completion of patient close-out	August 1974
VI. Termination	Termination of funding for original trial	March 1979
VII. Post-trial follow-up (optional)	Termination of all follow-up	December 1983

*The CDP is considered to have started in 1961 with the first planning meeting. The initial funding for the trial was awarded in March of 1965.

3.3 COMMON IMPEDIMENTS TO THE ORDERLY PERFORMANCE OF ACTIVITIES

3.3.1 Separation of responsibilities in government-initiated trials

The responsibilities for planning and executing a trial rest with the investigators in the typical investigator-initiated trial. They design it, they propose the investigators to be involved in it, and they carry it out. The sponsor has only a peripheral role. The situation is different in a typical sponsor-initiated trial. In this case, the sponsor assumes major responsibility for design of the trial and for selection of the investigators to carry it out. The separation of the design and execution functions may lead to sponsor-investigator tensions that may impede progress in the trial if they are not addressed.

3.3.2 Structural deficiencies

In a survey of multicenter trials, Smith (1978) classified over half of the operational problems encountered as organizational or administrative in nature. Many of these organizational problems can be traced to ambiguities in decision-making processes for resolving key design and operational issues (e.g., when to stop patient recruitment, how long to continue patient follow-up, when to terminate a treatment because of adverse or beneficial effects). The ambiguities can cause different individuals or groups to view themselves as the "final authority" in resolving a particular issue and can cause delays and inefficiencies in the way activities are conducted.

3.3.3 Overlap of activities from stage to stage

The activities normally associated with a particular stage may continue into the next or subsequent stages. Experience during the patient recruitment stage may require re-evaluation of sample size and other criteria set down when the study was designed. New treatments may be added after the start of patient recruitment, for example, as in the UGDP (University Group Diabetes Program Research Group, 1970d).

Similarly, it is rare for patient recruitment to be completed by the time treatment and follow-up begin. In fact, it is not uncommon for all three of these processes to go on simultaneously in long-term trials. In addition, data analyses, while typically associated with the termination stage, may be necessary long before that point is reached for performance and treatment monitoring, as discussed in Chapters 16 and 20, respectively.

Overlap of activities from one stage to the next has staffing implications. A trial in which patients are still being recruited, while others are in various stages of follow-up or have already been separated from the study, requires more elaborate organization and staffing than one in which it is possible to complete one stage before the next one starts.

3.3.4 Inadequate time for planning, development, and implementation

The time schedule for a trial, as established in the design stage, often proves to be unrealistic. Among the ten Requests for Proposals (RFPs) reviewed in the CCMP, only six made any mention of a time period for planning and protocol development (Coordinating Center Models Project Research Group, 1979b). The start-up time (i.e., time from start of funding to the enrollment of the first patient) for the trials listed in Appendix B ranged from 2 months to 3 years. The average time was just over 1 year.

Unrealistically ambitious time schedules tend to exert pressure on investigators to initiate data collection before the necessary data forms and related documents have been fully developed and tested. Doing so can lead to a chronic crisis atmosphere in the data center as staff struggle to develop better data forms and intake procedures while trying to maintain existing procedures.

3.3.5 Inadequate funding

The level of activities in a trial should be compatible with available funding. It is a mistake to embark on a trial without adequate support. The effort proposed should be scaled to match available support. Further, funds should be equitably distributed across activities within the trial. Situations should be avoided where support for one aspect of the trial, such as data collection, is overfunded, while another, such as data intake and analysis, is underfunded. A successful trial requires balance in the amount of money available for all essential activities.

3.4 APPROACHES TO ENSURE ORDERLY TRANSITION OF ACTIVITIES

3.4.1 Phased initiation of data intake

It may be prudent to limit the number of patients to be enrolled at the outset, especially if a clinic has a large backlog of patients waiting to be enrolled. The limit may be lifted once a clinic has demonstrated proficiency in the data collection process and after the basic data forms and intake procedures have been shown to work.

One approach to phased data collection in trials with multiple clinics involves funding only a small number of clinics at the outset, with new clinics being added as the trial proceeds. This approach was used in the CDP. It started with five clinics. Additional clinics were added over a 2-year period to make up the total of 55 ultimately involved in the trial (Coronary Drug Project Research Group, 1973a).

A gradual progression to full-scale recruitment and data collection can be part of the study plan, even if all the participating clinics are identified from the outset. It may be wise in such cases to designate one or two clinics to serve as testing sites for the treatment protocol and data collection procedures before the others are brought into the study. This approach was used in the Multiple Risk Factor Intervention Trial (Sherwin et al., 1981). Another approach allows all clinics to begin recruitment at the same time, but at a reduced rate to start with. The Hypertension Prevention Trial—HPT (see Sketch 13, Appendix B) used this approach. Each of the 4 clinics in that trial was required to enroll a test cohort of 20 patients before it was allowed to start full-scale recruitment.

3.4.2 An adequate organizational structure

Coordination of activities in a trial requires a sound organizational structure. One of the first orders of business should be its development. A sound structure takes time to develop and to reach maturity. There should be adequate time for that maturation process before the start of patient intake. As a rule, the period of time required for this process is related to the size and complexity of the trial, and it may be longer for sponsor-initiated trials than for investigator-initiated trials. A well-designed investigator-initiated trial will include details on organization in the funding application. The period of time between submission of the application and initiation of funding (see Section 21.2.1 of Chapter 21) may provide investigators with opportunities to refine the structure proposed and may even allow it to reach a degree of functional maturity because of investigator interactions required in preparing and defending the funding request. Such opportunities do not exist in the typical sponsor-initiated trial because of the way centers are selected (see Section 21.3 of Chapter 21).

3.4.3 Opportunities for design modifications in sponsor-initiated trials

The separation of responsibilities discussed in Section 3.3.1 is an inherent feature of most sponsor-initiated trials, especially those initiated by the government via RFPs. The timetable for the trial should provide investigators with adequate opportunity to consider and accept the design tenets proposed before the start of data collection. This process begins before the proposal is submitted in the typical investigator-initiated trial, but cannot begin until after the centers are selected and funded in the typical government-initiated trial.

3.4.4 Certification as a management tool

Patient recruitment should not start until the clinics and data center have demonstrated that they are properly staffed and equipped to support this activity. Some trials, such as the National Cooperative Gallstone Study (see Sketch 5, Appendix B), have required clinics to carry a minimum number of patients through key study procedures before recruitment could begin. A formal certification of clinics was required in the HPT prior to the start of recruitment.

The certification process has been extended to individuals making key measurements in some trials (e.g., see Early Treatment of Diabetic Retinopathy Research Group, 1982; Knatterud, 1981; Rand and Knatterud, 1980). The personnel certification process is useful in that it provides a landmark that must be passed before a person is cleared for data collection in a trial.

3.4.5 Realistic timetables

The timetables for activities proposed in grant applications or RFPs for clinical trials should be based on realistic appraisals of times required to complete those activities. Unrealistically ambitious schedules may raise doubts regarding the feasibility of the study in the minds of those responsible for overseeing it, may lead to frustration among investigators in the trial, and may result in decisions to implement activities before the required procedures and support systems have been adequately tested and developed. The timetable constructed at the beginning of a trial should be reviewed and, when necessary, revised as the trial proceeds if it is to retain its value as a management tool and performance monitoring standard over the course of the trial.

3.4.6 Ongoing planning and priority assessment

Planning and priority assessment are continuing needs in a trial. The leadership of the trial has a responsibility for implementing an active review process in order to make certain that work schedules and goals are compatible with the needs and resources of the trial. When they are not, priorities must be revised to reflect reality.

The leadership committee of the trial should take responsibility for setting priorities for data analyses when demands for them exceed resources available in the data center for carrying them out. The failure of the leadership committee to act in this capacity will leave staff in the data center open to criticisms if the priorities they set are not acceptable to everyone in the trial.

3.4.7 Minimal overlap of activities

The mix of activities under way at any one time influences the staffing needs of centers in the trial. The greater the heterogeneity of activities, the larger the staffing needs. The goal should be to minimize the number of activities under way at any one time. Pursuing this goal requires completion of patient recruitment in the shortest possible time. This means that all clinics in a multicenter trial should be prepared to continue patient enrollment until the study recruitment goal is met, even if some clinics exceed their goals while others fall short of theirs. For example, the CDP cut off patient enrollment at all clinics at the same time, even though it used a phased approach to clinic enrollment (see Section 3.4.1). Clinics that achieved their stated recruitment goal were asked to continue enrollment in order to reduce the time needed to achieve the study-wide recruitment goal of 8,300. Allowing each clinic to cut off recruitment when it achieves its prestated goal is inefficient for the data center, especially if there is wide variability among the clinics as to when the cut-off occurs. The data center will be required to maintain treatment allocation and baseline data intake procedures as long as recruitment continues in any clinic.

Similarly, the patient close-out process is most efficient when all patients are separated from the trial at the same time, regardless of when they were enrolled. The alternative is to separate each patient after a specified period of follow-up (e.g., 2 years). However, this approach is inefficient when patient recruitment has extended over a long period of time. See Chapter 15 for discussion.

4. Single-center versus multicenter trials

It is not the fault of our doctors that the medical service of the community, as at present provided for, is a murderous absurdity. . . . To give a surgeon a pecuniary interest in cutting off your leg, is enough to make one despair of political humanity. . . . And the more appalling the mutilation, the more the mutilator is paid. He who corrects the ingrowing toe-nail receives a few shillings; he who cuts your insides out receives hundreds of guineas, except when he does it to a poor person for practice.

George Bernard Shaw

4.1 DEFINITION

A center, in this book, is defined as any autonomous unit in a clinical trial that is involved in the collection, determination, classification, assessment, or analysis of data, or that provides logistical support for the trial. Included are clinical centers, data centers, coordinating centers, project offices, central laboratories, reading centers, quality control centers, and procurement and distribution centers. To qualify as a center, a unit must have a defined function to perform during one or more stages of a trial. In addition, it must be administratively distinct from other centers in the trial, and must be made up of two or more individuals who devote some portion of their time to the defined functions of the center.

A trial, to be considered as multicenter in this book, must involve:

- Two or more clinics
- A common treatment and data collection protocol
- A center to receive and process study data

All other trials will be considered single-center. This category includes:

- A single clinic, with or without satellite clinics (see Glossary) and with or without a center to receive and process study data or other resource centers (see Glossary)
- A trial involving multiple clinics, with or without satellite clinics, but not having a common study protocol, regardless of whether it has a center to receive and process study data
- A trial involving multiple clinics, with or without satellite clinics, that does not have a center to receive and process study data, even if clinics purport to follow a common study protocol
- A trial, such as the Physicians' Health Study (PHS), that does not involve any clinical centers, even if it has multiple resource centers

The four elements of the definition are necessary with the binary language structure used to characterize the physical structure of trials. However, the fact is that most trials are characterized by the first element in the category and, hence, they are discussed from this perspective throughout this book.

4.2 NATIONAL INSTITUTES OF HEALTH (NIH) COUNT OF SINGLE-CENTER AND MULTICENTER TRIALS

The 1979 NIH Inventory of Clinical Trials was the first inventory generated by that agency that distinguished between single-center and multicenter trials (National Institutes of Health, 1975, 1980). The institutes vary widely with regard to support for the two types of trials.[1] For example, all of the 26 trials supported by the National Institute of Dental Research were single-center, whereas all but 1 of the 20 trials sponsored by the National Heart, Lung, and Blood Institute were multicenter (Table 4–1). The differences are due, in part, to the nature of the evaluation question faced by the various institutes (see Section 2.1).

Overall, the institutes of the NIH sponsor about as many multicenter trials, 476, as single-center trials, 510 (last line, Table 4–1). It is interesting, in view of this fact, to note the preponderance of single-center trials in published literature. Only 25% of the 306 gastrointestinal trials reviewed by Juhl and co-workers (1977) involved multiple clinics. Chalmers and co-workers (1972), in their review of cancer trials identified only 49 as multicenter trials out of 2?? reviewed. Only 15 of the 113 trials published in 1980 and reviewed for this book were multicenter by the definition used in this book (Table 4–3).

4.3 DESIGN CHARACTERISTICS OF SINGLE-CENTER VERSUS MULTICENTER TRIALS

Table 4–2 provides a summary of a few of the key design features of single-center trials versus multicenter trials for NIH-sponsored trials reported in the 1979 NIH Inventory (National Institutes of Health, 1980). Table 4–3 provides a corresponding summary for the 113 trials discussed in Chapter 2.

A major difference between multicenter and single-center trials, apparent in both tables, is sample size. The typical multicenter trial has more patients than does the typical single-center trial. This difference is most apparent for the 113 papers reviewed in Chapter 2. The median number of patients enrolled per trial was 283 for the 15 multicenter trials, which contrasts with 40 for the 98 single-center trials (Table 4–3).

1. The definition of multicenter trials used by the NIH is less stringent than the one stated above. Trials in the Inventory were classified as multicenter without the requirement of a common protocol or the presence of a center to receive and process study data.

Table 4–1 NIH-sponsored single-center and multicenter trials by institute for fiscal year 1979

Sponsoring institute	*Total number of trials*	*Single-center*		*Multicenter*	
		Number	*Percent*	*Number*	*Percent*
National Cancer Institute (NCI)	654	261	39.9	393	60.1
National Eye Institute (NEI)	26	18	69.2	8	30.8
National Heart, Lung, and Blood Institute (NHLBI)	20	1	5.0	19	95.0
National Institute of Allergy and Infectious Disease (NIAID)	120	104	86.7	16	13.3
National Institute of Arthritis, Diabetes, and Digestive and Kidney Diseases (NIADDK)	67	42	62.7	25	37.3
National Institute of Child Health and Human Development (NICHD)	32	29	90.6	3	9.4
National Institute of Dental Research (NIDR)	26	26	100.0	0	0.0
National Institute of Neurological and Communicative Disorders and Stroke (NINCDS)	40	29	72.5	11	27.5
National Institute of General Medical Sciences (NIGMS)	1	0	0.0	1	100.0
Total	986	510	51.7	476	48.3

Table 4–2 Design features of NIH single-center and multicenter trials

Feature	*Single-center*		*Multicenter*	
	Number	*Percent*	*Number*	*Percent*
Total number of trials	510	100.0	476	100.0
Number of treatment groups/trial				
1	159	31.2	99	20.8
2	217	42.6	221	46.4
≥3	134	36.3	156	32.8
Median number	2		2	
Sample size				
Median number of patients/trial	60		166	
Range*	25 to 200		52 to 362	
Number of patients/trial/treatment group†	30		83	
Range*	12 to 100		26 to 181	
Method of treatment allocation				
Random	259	50.8	334	70.2
Nonrandom	251	49.2	142	29.8
Type of trial				
Therapeutic	369	72.5	432	90.8
Prophylactic	90	17.7	36	7.6
Diagnostic	50	9.8	8	1.7
Anticipated length of funding				
>1 year ≤ 2 years	10	2.0	9	1.9
>2 years ≤ 3 years	51	10.0	50	10.5
	129	25.3	94	19.7
	319	62.7	323	67.9

*20th to 80th percentile.
†Calculated by dividing median number of patients per trial by the median number of treatment groups per trial.

4.4 THE PROS AND CONS OF SINGLE-CENTER VERSUS MULTICENTER TRIALS

Certain features of single-center trials make them appealing. They are generally easier to mount and carry out than their multicenter counterparts. The fact that all study personnel are located in the same institution in most single-center trials obviates the need for and expense of maintaining communications and decision-making structures needed for execution of most multicenter trials. In addition, the physical proximity of study personnel may make it possible for them to work more efficiently and to achieve a higher degree of uniformity in the procedures they perform than might be expected in a multicenter trial. Further, the fact that all patients enrolled in the trial come from the same area in the typical single-center trial should produce a more homogeneous study population than might be expected of a population made up of patients from different clinics.

The main weaknesses of the single-center trial are sample size and resource limitations. One center and a few investigators will find it difficult to recruit and follow the numbers of patients needed. Compromises will have to be made in order to bring the number of patients required for study into line with reality while still providing adequate type I and II error (see Glossary) protection. The original trial, planned to focus on a single clinical event as the outcome, may have to be converted to one involving composite forts involved in mounting and carrying out a multicenter trial. It is much easier and less time consuming to design and carry out a short-term trial in a single clinic than it is to mount and execute one extending over a period of years and involving multiple clinics. Most investigators lack the time and wherewithal to initiate such trials. And even if they do have the resolve to carry such efforts forward, they may not have the support needed to cover developmental costs for the work. The demise of NIH planning grants has virtually precluded the acquisition of government funds for planning multicenter trials. As a result, responsibility for initiative rests in the hands of senior investigators with other sources of support and in the hands of sponsoring agencies.

Another reason for the prominence of single-center trials is that promotions in most academic institutions are based, in large measure, on the originality, number, and quality of papers produced by those considered for promotion. As a result, an investigator who carries out a number of short-term, single-center trials and who uses them to produce a series of papers as sole or senior author is more likely to be promoted than one who works on a few long-term multicenter trials and who produces relatively few papers, even if of high quality. The prospects for promotion may be further diminished if the papers produced are written under a corporate masthead (see Remington, 1979, and see also Chapter 24 for a discussion of authorship policies).

4.7 TIMING OF SINGLE-CENTER VERSUS MULTICENTER TRIALS

Many investigations of a new or existing treatment modality begin with uncontrolled observational studies, followed by small-scale clinical trials. Only after the results of these trials begin to appear in print, and especially if they are inconclusive or conflicting, is the need for larger trials recognized. Even then, sponsors and the review groups that advise them will be reluctant to commit the money required for a multicenter trial if they think answers can be obtained with less effort and money.

Some evaluation questions are slow to progress beyond the stage of uncontrolled studies; some never progress beyond that point. Others may be considered only in the context of multicenter trials from the outset. A case in point is risk factor reduction for cardiovascular disease. There is no realistic way to address this issue except via large-scale trials, such as MRFIT (Multiple Risk Factor Intervention Trial Research Group, 1982).

Three general conditions should be satisfied before a multicenter trial is considered. First, there should be evidence that multiple clinics are needed to meet the sample size requirements of the trial. A single-center trial may suffice if the sample size requirement is modest. Second, there should be an identifiable group of clinical investigators who are willing and able to follow a common treatment and data collection proto-

Table 4-5 NIH expenditures for trials in fiscal year 1979 by type of trial

	Trials		*Amount (millions of dollars)*		
Type of trial	*Number*	*Percent*	*Dollars*	*Percent*	*Median patient cost per year**
Single-center	510	51.7	35.0	25.7	$587
Multicenter	476	48.3	101.1	74.2	$523
Total	986	100.0	136.2	100.0	$574

*The dollar cost per patient per year for a given trial was derived by dividing the total projected expenditures for that trial by the product of the number of patients to be enrolled (projected) and years of support (projected) required for execution of the trial. The median dollar cost per patient per year for a given type of trial was determined by ranking the resulting figures for individual trials from lowest to highest and then locating the dollar value corresponding to the 50th percentile point in the resulting distribution (median value).

col. Third, there should be an identifiable set of [illegible] with adequate support staff and facilities to carry out the trial.

4.8 COST OF SINGLE-CENTER VERSUS MULTICENTER TRIALS

The only database available for a comparative analysis of cost is that provided via the 1979 NIH Inventory of Clinical Trials.[2] The total dollar cost for multicenter trials was nearly three times that for single center trials in 1979 (101.1 million versus 35.0 million). However, this figure is misleading in that it is not adjusted for the differences in sample size noted in Table 4-2 for the two types of trials. This has been done in Table 4-5 using median cost per patient per year of study. When viewed in this way, the cost is actually less than for single-center trials—a noteworthy fact in view of oft-expressed concerns regarding the cost of multicenter trials.

[2] See footnote 1, page 12.

5. Coordinating and other resource centers in multicenter trials

Technical skills, like fire, can be an admirable servant and a dangerous master.
A. Bradford Hill

5.1 INTRODUCTION

A resource center is any center involved in a trial, other than a clinical center, that is in charge of performing a specific set of functions concerned with the design, conduct, or analysis of the trial. Resource centers include (see Glossary for definitions):

- Data centers
- Data coordinating centers
- Treatment coordinating centers
- Coordinating centers
- Project offices
- Central laboratories
- Reading centers
- Quality control centers
- Procurement and distribution centers

This chapter focuses on coordinating centers because of their key role in the typical multicenter trial. The coordinating center, or data coordinating center when there are separate coordinating centers for data collection and treatment, will be among the first to be funded and the last to cease operations when the trial is completed. It may, in fact, operate after the trial is terminated if post-trial follow-up (see Glossary) is required.

All 14 trials sketched in Appendix B included either a coordinating center or data coordinating center. No other resource center was common to all the trials (Table 5-1).

5.2 COORDINATING CENTERS

As noted in the previous chapter, a multicenter trial is defined herein to include a center that is

Table 5-1 Type of resource center represented in the 14 trials sketched in Appendix B*

Type of center	Number of trials with center†
Coordinating center‡	14
Project office	13
Reading center	12
Central laboratory	11
Procurement and distribution center	6
Quality control center	2

* All of the trials were classified as multicenter except one, the Physicians' Health Study (PHS).
† Several trials had multiple laboratories and/or reading centers. See Table B-4, Appendix B for specifics.
‡ Some of the trials had both a data and treatment coordinating center.

responsible for receiving, editing, processing, analyzing, and storing data generated in the trial. In fact, some studies may use multiple centers to perform this function. The most common approach when this is the case, is to establish regional data centers, with each of the centers performing identical functions. Such structures, while relatively uncommon for studies done in one country, may be necessary in international studies, especially when different languages are involved. Both the International Reflux Study in Children, IRSC (see Sketch 14, Appendix B), and the International Mexiletine Placebo Antiarrhythmic Coronary Trial, IMPACT (Alamercery et al., 1982) had separate data coordinating centers to service United States and European based clinics.

The data center (or centers), at least in the larger multicenter trials, will typically have a number of coordination responsibilities. This book makes a distinction between two types of coordinating functions—those related to data collection and those related to treatment. A *data coordinating center* is defined as one that, in addition to responsibilities for receiving, editing, processing, analyzing, and storing data generated in a trial, has responsibilities for coordinating the data generation activities of the clinics and for implementing and maintaining quality assurance procedures related to the data generation process. Responsibilities for coordinating the administration of treatments in the trial and the surveillance of clinic activities are vested in a second center—a *treatment coordinating center*. The unmodified term *coordinating center* will be used to designate a center that fulfills both the data and treatment coordination functions.

Use of the term *coordinating center* outside this book does not always conform to these conventions. For example, the facility designated as the coordinating center in the National Cooperative Gallstone Study (NCGS) was responsible for treatment coordination and for dispersal of funds to the other participating centers, but had no data coordinating responsibilities. The center with those responsibilities in the NCGS was referred to as the Biostatistical Center (National Cooperative Gallstone Study Group, 1981a).

5.2.1 General activities

The general activities of the coordinating center by stage of the trial are summarized in Table 5-2 (see also Appendix D). The list is adapted from one developed in the Coordinating Center Models Project, CCMP (Coordinating Center Models Project Research Group, 1979a, 1979d). The activities listed for the first stage—the initial design stage—and some of those for the second stage—the protocol development stage—may be assumed by the sponsor in sponsor-initiated trials.

No one center will necessarily have responsibilities for all the functions listed, especially if there are separate centers for treatment and data coordination. A review of coordinating centers for the trials included in the CCMP revealed important differences in their duties, partly because of the differences in the roles assumed by other units in the trial, most notably the project office and the office of the study chairman (McDill, 1979).

One of the major responsibilities of the coordinating center relates to preparation and distribution of key study documents, such as the manual of operations and data collection forms. In addition, the center typically serves as the repository for completed data forms (except for studies with distributed data entry systems), minutes of study meetings, progress reports, performance monitoring reports, and treatment effects monitoring reports.

5.2.2 Location

The coordinating center, under ideal circumstances, will be administratively and physically distinct from the sponsor and from all other cen-

Table 5-2 Coordinating center activities by stage of trial, with emphasis on data coordination activities

Initial design stage

- Calculate required sample size
- Outline data collection schedule, quality control procedures, data analysis plans, and data intake and editing procedures
- Develop organizational structure of the trial
- Prepare funding proposal for coordinating center
- Coordinate preparation of the funding application

Protocol development stage

- Develop treatment allocation procedures
- Develop computer programs and related procedures for receiving, processing, editing, and analyzing study data
- Design and test data forms
- Develop interface for data transmission from clinics and other resource centers to coordinating center
- Train clinic personnel in required data collection procedures
- Implement clinic and personnel certification procedures
- Distribute study data forms and related materials
- Develop manuals needed in the trial, including the treatment protocol, clinic manual of operations, coordinating center manual of operations, etc.
- Provide a repository for official records of the study, including minutes of meetings, manuals of operation, etc.
- Serve as the funding center for a trial operated under a consortium agreement, unless this function is fulfilled by some other center
- Serve as the payment center for general study needs, such as study insurance, and other specialized procedures not provided for in the grants or contracts of other participating centers

Patient recruitment stage

- Administer treatment allocations, including checks for breakdowns in the assignment process
- Assume leadership role in outlining study needs for quality assurance
- Implement editing procedures to detect data deficiencies
- Develop performance monitoring procedures and prepare data reports to summarize performance of participating clinics
- Develop treatment monitoring and reporting procedures to detect evidence of adverse or beneficial treatment effects
- Respond to requests for analyses from within the study structure
- Site visit participating clinics
- Prepare study progress reports for submission to sponsor
- Prepare, in conjunction with the study leadership, renewal or supplemental funding requests
- Update study manuals

Treatment and follow-up stage

- Prepare periodic data reports for safety monitoring committee
- Prepare periodic reports on performance of clinics and resource centers
- Carry out periodic training sessions to maintain a high level of proficiency at clinics in treatment and data collection procedures
- Evaluate data processing procedures and modify as necessary
- Develop and test data collection forms for close-out stage
- Prepare summary of study results for presentation to participating investigators for use in close-out stage
- Assume responsibility for location of patients lost to follow-up
- Take initiative for reviewing study priorities and for proposing changes in the organizational or operating structure of the trial
- Assume major role in writing paper on design and methods

Patient close-out stage

- Monitor for adherence to agreed-upon patient close-out procedures
- Develop plans for final data editing
- Design and test computer programs needed for final data analysis
- Develop plans for final disposition of study data
- Coordinate logistics of patient disengagement from treatment
- Assume key role in writing papers summarizing results of the trial
- Develop plans for disengagement of clinical centers from the trial

Termination stage

- Perform final data edit and undertake final analyses of data according to plans outlined by study leadership
- Implement study plans for disposition of study records
- Assume leadership role in paper writing activities
- Undertake extra measures to locate patients lost to follow-up
- Supervise collection and disposal of unused study medications
- Distribute draft manuscripts and published papers to participating centers
- Serve as funding center for activities in the trial after termination of support for clinics

Table 5-2 Coordinating center activities by stage of trial, with emphasis on data coordination activities (*continued*)

Post-trial follow-up stage (optional)

- Compile a list of patients eligible for post-trial follow-up
- Implement procedures to locate patients whose current whereabouts are unknown
- Coordinate mailings, telephone calls, or clinic visits required for post-trial follow-up
- Update existing data files with data collected during post-trial follow-up
- Assume leadership role in drafting and distributing any manuscript using post-trial follow-up results
- Store, under adequate security, names of study patients and other identifying information for future follow-up

centers in the trial. This separation insulates the center from the direct administrative control of the sponsor, and helps it to establish and maintain balanced working relationships with all other centers in the trial. This balance may be difficult to achieve if the center is part of the sponsoring agency or if it is physically or fiscally a part of one of the clinics in the trial.

Twelve of the 14 trials sketched in Appendix B had coordinating or data coordinating centers located in academic institutions. This setting has pluses and minuses. A prestigious teaching institution, especially one with a recognized degree program in biostatistics, epidemiology, or related fields, provides a pool of bright and energetic people to meet the programming and data analysis needs of the center. In addition, the opportunity to teach and to interact with other faculty may help the center attract and retain senior professional personnel.

The minuses stem from the internal bureaucracy of any large academic institution. Most of the coordinating centers reviewed in the CCMP (all except one of ten centers reviewed were located in academic institutions) complained of difficulties in recruiting intermediate-level personnel because of pay and promotional restrictions imposed by their respective institutions. Several had difficulty in purchasing computing hardware for their own needs because of policies aimed at discouraging dedicated facilities.

The real or perceived lack of administrative flexibility of such settings, coupled with small business set-asides for government-funded studies (United States Congress, 1981), has given impetus to coordinating centers located in private (profit or nonprofit) business firms. The Persantine Aspirin Reinfarction Study (PARIS) coordinating center, located at the Maryland Medical Research Institute, is a case in point (Persantine Aspirin Reinfarction Study Research Group, 1980a). The main advantage of this setting is the administrative flexibility it provides for personnel hiring and pay practices and for acquisition of needed computing hardware and software. The main disadvantage stems from the lack of stability of any operation devoted to a specialized set of activities. That lack may make it difficult to recruit and retain needed personnel.

5.2.3 Staffing

Ten of the 14 trials sketched in Appendix B had coordinating centers headed by persons with a doctorate in biostatistics. Three centers were headed by persons with M.D. degrees; and one was headed by a person with a master's degree in applied mathematics.

All coordinating centers require expertise in the areas of biostatistics and computer programming. Ideally, the staff should include someone trained in medicine who is knowledgeable in the disease under treatment as well. When this is not possible, the director of the center should establish a working relationship with appropriate medical personnel located outside the center. The relationship may be established via collaboration with a medical department in the director's parent institution or nearby medical facility, or via relations with one of the clinics in the trial.

The CCMP has provided summary staffing data for seven of the coordinating centers reviewed in that project (Hawkins, 1979). A detailed staffing profile for the Coronary Drug Project (CDP) coordinating center is provided in Table 5-3 (see also Meinert et al., 1983). The figures in the table were based on data contained in annual budget requests of the CDP coordinating center to the National Heart, Lung, and Blood Institute (NHLBI).

The total number of full-time equivalents (FTEs) rose from 7 in the first year to a high of 36 in the tenth year (column 3 of Table 5-3). Programmers and master's-level statisticians accounted for about one-quarter of the staff during the 13-year period covered in the table

Table 5-3 Percent of full-time equivalents by category of personnel and year of study for the CDP Coordinating Center

Year of study*	Stage	Total FTEs	MD or PhD in statistics	MSc in statistics	Data coords, key punch, coders	Support personnel†
			Percent of full-time equivalents (FTEs)			
1st	Protocol dev.	6.8	27.0	29.2	29.2	14.6
2nd	Recruitment	14.8	32.4	20.3	27.0	20.3
3rd	Recruitment	19.0	20.1	31.5	31.5	16.8
4th	Recruitment	24.3	19.8	28.8	32.9	18.5
5th	Follow-up	24.8	17.3	24.2	32.3	26.2
6th	Follow-up	26.4	18.6	22.7	26.5	32.2
7th	Follow-up	27.3	17.6	22.0	25.6	34.8
8th	Follow-up	30.3	15.8	26.4	23.1	34.6
9th	Follow-up	29.5	16.3	23.7	24.4	35.6
10th	Close-out	35.7	12.3	23.8	23.8	40.1
11th	Termination	22.6	14.2	24.8	23.5	37.6
12th	Termination	17.2	18.6	27.9	19.2	34.3
13th	Termination	11.5	21.7	25.2	17.4	35.7

Source: Reference citation 320. Reprinted with permission of Elsevier Science Publishing Co., Inc., New York.

*The study started in April 1965. Patient recruitment began near the end of the first year (March 1966) and was completed during the fourth year of the study (October 1969). Close-out of follow-up occurred in 1974 during the first half of the tenth year. The main activity thereafter had to do with analyses for paper-writing activities.

†Administrative, secretarial, and clerical personnel. Also includes a graphic artist.

(column 5). Data processing activities were concentrated on systems development and programming for data intake and editing during the early part of the trial. Reductions in these activities as the study progressed were offset by increased demands for data analyses.

Data coordinators, key-punch operators, and coders accounted for one-quarter to one-third of all coordinating center personnel through the eleventh year (column 6 of Table 5-3). The drop in years 12 and 13 resulted from reductions in data intake and keying operations following completion of patient close-out.

Secretarial, clerical, and administrative staff constituted the largest personnel category starting with the sixth year. Growth of this category from 15% in the first year to more than 40% of the FTEs in the tenth year was a reflection of an increasing workload associated with manuscript production and maintenance of various reading and quality control procedures in the study.

5.2.4 Equipment

The equipment listed in Table 5-4 represents items that are likely to be needed in a "typical" coordinating center operation. The list does not include general office furniture and equipment, such as desks, chairs, typewriters, and dictating and transcribing equipment. These are assumed to be part of any office setting.

The approach a coordinating center takes to data entry and processing may be dictated in large measure by the equipment that exists at the institution housing the center and the data processing philosophy held by key people in that institution. The factors that should be considered in choosing between a dedicated or centralized approach to computing are discussed in Chapter 17.

5.2.5 Relative cost

Table 5-5 provides data on the relative cost of five coordinating centers reviewed in the CCMP (Meinert, 1979a). The percentage annual cost of the individual centers, relative to the total cost of the trials for that year, ranged from 5.1 to 51.7 in the first year and from 7.3 to 16.8 for the other years covered in the table.

Figure 5-1 is based on data from the CDP (Meinert et al., 1983). As in Table 5-5, the values reported represent the proportionate cost of the coordinating center, expressed as a percentage of the total direct cost of the study. The major share of the expenditures during the first year occurred in connection with equipment purchases

Table 5-4 General equipment requirements of coordinating centers

- Computing facilities* for storing, editing, and analyzing study data
- CRT work stations for use by programming and data processing staff
- RJE station with high-speed printer
- Dedicated minicomputer for data storage and simple analyses
- Computer-controlled graphics equipment
- Electronic calculator*
- Data entry equipment* (e.g., key punches, key-to-tape units, key-to-diskette units)
- Word processing equipment*
- Photocopying equipment*
- Collation and report binding equipment
- Telecopier (for transmitting and receiving special documents)
- Mailing equipment (postage meter, scale, etc.)
- Filing cabinets with locks*
- Microfilming equipment and viewers
- Fireproof, environment-controlled storage vaults for data tapes and other essential study documents

*Essential items

and work in developing data forms and manuals for the study. Expenditures in the clinics were modest until the start of patient recruitment in the second year. Support for clinics terminated during the eleventh year. Only the coordinating center was supported beyond that time. The gradual increase in proportionate costs starting with the third year and continuing through the tenth year is a reflection of increased demands for analyses related to treatment and performance monitoring and for paper writing, superimposed on continuing demands for maintenance of established data collection, intake, and editing procedures.

There are no accepted rules of thumb for determining the correct allocation of funds for the coordinating center, relative to other centers in the trial. The amount will depend on the nature and complexity of the data collection, editing, and analysis procedures needed, and on the total number of clinical centers in the trial. The relative costs, all other things being equal, will fall as the number of clinics increases, since many of the developmental, programming, and analysis costs incurred by the coordinating center are independent of the number of clinics. Part of the drop in relative cost, shown in Figure 5-1, is due to the addition of new clinics during the first two years of the CDP. There were only five clinics funded during the first year. Twenty-three additional clinics were funded early in the second year. The last complement of 27 clinics was added near the end of the second year.

The funds available for the coordinating center must be in line with the demands placed on it. Experienced investigators and sponsors will review the overall allocation of funds at intervals over the course of the trial and will reallocate funds among centers if there are gross imbalances. The way in which this is done depends on the funding vehicle. It is relatively easy to do with either a consortium approach to funding or with contracts, but not when each center has its own grant (see Chapter 21).

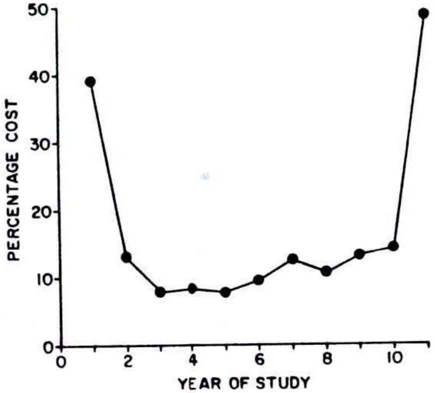
Table 5-5 Relative cost of coordinating centers for five trials* reviewed in the Coordinating Center Models Project

Year of trial	Number of trials†	Lowest	Median	Highest
		Percent of total study cost		
1	5	5.1	9.0	51.7
2	5	7.3	9.7	16.8
3	5	8.6	10.1	14.0
4	4	9.7	10.1	13.6
5	4	9.8	11.2	13.6
6	3	10.7	11.4	16.1

*Aspirin Myocardial Infarction Study (AMIS), Coronary Drug Project (CDP), Hypertension Detection and Follow-Up Program (HDFP), Lipid Research Clinics, Coronary Primary Prevention Trial (LRC-CPPT), and Multiple Risk Factor Intervention Trial (MRFIT).

†AMIS reported data through the third year. MRFIT reported data through the fifth year.

Figure 5-1 Percentage cost of the CDP Coordinating Center, relative to total direct study cost.*



*Based on direct cost expenditure data from the NHLBI, excluding costs for the central laboratory and drug distribution center. Total costs for all the centers combined ranged from $3.3 to $4.3 million during the third through the tenth year of the study. Figures for the first two years and the eleventh year were $0.5, $1.3, and $0.9 million, respectively.

Source: Reference citation 320. Reprinted with permission of Elsevier Science Publishing Co., Inc., New York.

5.2.6 Internal allocation of funds

The allocation of funds within the coordinating center is as important as the allocation of funds among centers. The amount of support available for personnel must be balanced against that available for equipment, computing, and other support services.

The internal allocation of funds, as reflected by annual budget requests submitted to the sponsoring agency for several different centers, is given in Table 5-6 (Meinert, 1979a). Table 5-7 provides a detailed look at the allocation of funds within the CDP coordinating center (Meinert, 1983). Ideally, the results in both tables should be based on after-the-fact expenditure data, but reliable data of this sort are almost impossible to obtain.

The typical coordinating center, as reflected by the median values recorded in Table 5-6, budgeted somewhere between 50 and 60% of its direct cost funds to personnel and about 20% to computing. The latter category includes funds for rental of data processing equipment, as well as time charges for computer use and for software rentals or purchases.

Funds requested for travel ranged from 3 to 6% of the annual budget. They were used to cover travel for center staff to attend study committee meetings, meetings of the entire investigative group, visits to participating centers, and scientific meetings. The "All other categories" in Tables 5-6 and 5-7 contain cost items needed to support general activities in the trials and include funds for items such as study publications, study insurance, and consultant fees and related expenses.

5.3 CENTRAL LABORATORIES

An issue in any trial that requires laboratory determinations is where those determinations are to be made. In this regard it is important to

Table 5-6 Budget allocation for coordinating centers by category and year of study. Results for centers from AMIS, CDP, CAST, HDFP, LRC-CPPT, and MRFIT

Year of study	Number of centers	*Median percent of direct costs devoted to:* Personnel	Computing†	Travel	All other categories
1	6	50	19	6	25
2	6	62	19	6	13
3	6	60	16	4	20
4	6	61	18	3	18
5	5*	60	18	4	18
6	4*	60	16	4	20

*Budget data were available for all six centers only through the first four years at the time the table was prepared. AMIS did not yield data for years 5 and 6. CASS did not yield data for year 6. Also, one center did not provide a personnel budget for year 5. Hence, the median value for personnel for that year is based on results from only four centers. All other entries for that year are based on five studies.

†Includes funds for computer time, as well as for purchase or rental of data entry equipment and for computing hardware and software.

Table 5-7 Budget allocation of the CDP Coordinating Center, by category and year of study

Year study	Stage	CC funds requested (direct costs)	*Percent of direct costs devoted to:* Personnel	Computing*	Travel	All other categories
1st	Protocol dev.	$188,111	29.0	55.7	2.1	13.2
2nd	Recruitment	196,103	75.7	8.9	4.1	11.3
3rd	Recruitment	279,749	73.0	12.2	2.9	11.9
4th	Recruitment	316,384	65.6	22.4	2.5	9.5
5th	Follow-up	372,242	68.1	22.1	2.4	7.4
6th	Follow-up	403,991	67.6	17.6	2.2	12.6
7th	Follow-up	507,745	75.1	15.4	2.1	7.4
8th	Follow-up	432,996	73.7	12.1	1.8	12.4
9th	Follow-up	569,170	72.5	16.9	1.9	8.7
10th	Close-out	595,756	73.8	17.1	1.8	7.3
11th	Termination	498,494	67.8	20.6	2.2	9.4
12th	Termination	396,023	64.2	24.7	2.5	8.6
13th	Termination	339,736	60.2	25.3	2.4	12.1

Source: Reference citation 320. Reprinted with permission of Elsevier Science Publishing Co., Inc., New York.

*Includes funds for computer time as well as for purchase or rental of data entry equipment and for computer hardward and software.

distinguish between determinations required for routine patient care and those needed for treatment comparisons. The former set of determinations may be performed locally and need not even be part of the central data file. The latter set of determinations may be done locally or in a central laboratory and should be part of the central data file.

All but three of the trials listed in Appendix B relied on central laboratories for making certain determinations. However, many of those same trials also relied on local laboratories for other determinations.

It will be necessary to rely on local determinations where it is impractical to use a central laboratory or where rapid feedback is required (e.g., in determining patient eligibility or in making treatment decisions that depend on laboratory values). Even if this is done, however, the determinations may be repeated at a central laboratory in order to provide results that are free of laboratory variation. In such cases, investigators must decide which set of determinations are to be used for assessment of patient eligibility and for treatment decisions.

The general factors to be considered in deciding whether to use a central laboratory at all are outlined in Table 5-8. The costs and logistical difficulties of establishing and operating a central laboratory must be balanced against need. Valid treatment comparisons can be made with results obtained from local laboratories as long

Table 5-8 Central versus local laboratories in multicenter trials

Local laboratory needed or permissible when:

- Specimens cannot be preserved for shipment to a central laboratory
- Determinations are needed quickly for the acute management of patients
- Higher level of precision possible through use of a central laboratory not essential to the trial
- All participating clinics have laboratories that perform the required determinations
- Individual laboratories are all certified by the same agency and are part of an ongoing standardization and quality assurance program
- Local laboratories agree to participate in standardization and monitoring efforts required by the study
- Senior personnel of each local laboratory are sensitive to the specific needs of the study and are willing to make adjustments in their procedures
- Risks of treatment feedback bias (i.e., where the laboratory reading obtained is influenced by knowledge of a patient's treatment) is minimal, e.g., as in double-masked trials

Central laboratory needed or desired when:

- Required determinations cannot be performed at the local laboratory
- Required level of standardization is not feasible with individual laboratories
- Separation of laboratory and clinics is needed to restrict flow of laboratory results back to clinics
- Laboratory measure is subject to wide variability from laboratory to laboratory

as the treatment allocations are balanced by clinic.

The fact that the central laboratory is remote from the clinics has advantages and disadvantages. The location adds to the cost and logistical difficulties involved in transport of the specimens. However, it also helps to ensure that the required masks are maintained (e.g., that the determinations are performed by personnel with no knowledge of patient treatments).

5.4 READING CENTERS

A reading center is a facility designed to provide the technical skills needed to read and code materials or records collected in the trial. The readings should be made by individuals who have no knowledge of the treatment assignment to ensure separation of the treatment and reading processes. They may involve extracting information from ECGs, fundus photographs of the eye, angiograms of the vascular structure of the heart, cholecystograms, chest x-rays, liver biopsies, food records, death certificates, or autopsy material. Of the 14 trials sketched in Appendix B, 12 had one or more such centers.

The conditions under which a centralized approach to reading is advantageous are outlined in Table 5-9. They are in large measure similar to those discussed for central laboratories.

The way in which central readings for eligibility assessments are to be used poses problems when they do not agree with local readings, if local readings are used for decisions on enrollment and randomization. Decisions must be made in such cases as to the disposition of patients where there are disagreements. Patients should be retained if the disagreements are minor. Procedures that allow investigators to exclude patients after randomization must be administered by personnel masked to treatment assignment and treatment results (see questions 38b, 39, and 50, Chapter 19).

The number of independent readings per record is a design question that should be resolved before any records are read. It is common to require two independent readings, with or without subsequent adjudication of disagreements. Duplicate readings offer a more precise basis for treatment comparisons than is possible with a single reading. However, valid comparisons can be made with just one reading per record, so long as the readings are independent of treatment assignment.

Table 5-9 Conditions under which centralized reading may be required

- Reading procedures are complex and require special skills or training
- High degree of uniformity and standardization required in the readings, especially for determining eligibility for the trial and for key items of follow-up information
- Large volumes of records are to be read
- Separation of the reading and treatment processes desired

5.5 PROJECT OFFICES

The project office, as defined in this book, is located at the sponsoring agency and is designed to serve as an interface between the sponsor and the investigative group involved in the trial. The main functions assumed by staff in the project office are to:

- Represent the interests of the sponsor in the design and operation of the trial
- Perform coordinating functions assigned by the leadership committee of the study
- Perform special functions assumed or assigned to the office by the sponsor or investigative group
- Serve as members of the key leadership committees of the study
- Carry out special analyses and tabulations

The National Institutes of Health (NIH) has used different terms to designate the office fulfilling these functions. It is usually designated as the project office but may have other names such as medical liaison office or program office. The role of the project office will be related to the perceived importance of the trial by the sponsoring agency and the size of its financial investment. Generally, the greater the investment, the greater the involvement of the project office. Its role will also be influenced by the responsibility of the sponsor in initiating the project. It will tend to have a more pronounced role in sponsor-initiated trials than in investigator-initiated trials.

There should be a well-defined division of responsibilities between the project office and the coordinating center. Failure to specify a division can lead to friction between the office and the center. Any division is workable so long as the principals involved understand and accept it.

The role assumed by the project officer is influenced by his or her personality. A strong, assertive person will automatically have an active role in the trial. The project officer's role is also heavily influenced by the personalities of others in the trial. The opportunity for an active role will be encouraged by a weak study leadership structure and discouraged by a strong one.

5.6 OTHER RESOURCE CENTERS

Several trials sketched in Appendix B included a drug procurement and distribution center. The VA Cooperative Studies Program has a general center located in Albuquerque, New Mexico, that fulfills this function for all its drug trials (Hagans, 1974; Veterans Administration Cooperative Studies Program, 1982).

PARIS had a quality control center. Its duties are outlined in one of the publications from that study (Persantine Aspirin Reinfarction Study Research Group, 1980a; see also Sketch 8, Appendix B). One of the prime functions of the center was to check on the accuracy of the data entry and analysis procedures carried out by the coordinating center. It also played a role in the development of new data analysis procedures for the trial.

The HPT (Sketch 13, Appendix B) includes a treatment coordinating center. One of its duties is to compile materials used in counseling study patients to make the required diet changes.

6. Cost and related issues

A man may do research for the fun of doing it but he cannot expect to be supported for the fun of doing it.

J. H[illegible]

6.1 GOVERNMENT EXPENDITURES FOR CLINICAL TRIALS

Table 6-1 gives a count of trials for the various institutes of the NIH by fiscal year (National Institutes of Health, 1975, 1980). The number of trials reported ranged from a low of 746 in fiscal year (FY) 1977 to a high of 986 in FY 1979. Table 6-2 gives the NIH expenditures for clinical trials as a percentage of total NIH appropriations. The dollar figures given for total appropriations are from an NIH fact book (National Institutes of Health, 1981a). Expenditures for clinical trials represented from 4.1 to 5.3% of total appropriations over the 5-year period covered in the table. (See Section 2.1 for notes on how the inventories were compiled.)

Table 6-3 gives expenditures by institute and fiscal year for clinical trials as a percentage of total NIH expenditures. The relative distribution of expenditures among institutes has remained fairly constant over the 5-year period covered. The National Heart, Lung, and Blood Institute (NHLBI) has had the largest expenditures for trials, even though the number of trials (Table 6-1) is small relative to some of the other institutes. This Institute plus the Cancer Institute accounted for over three-fourths of all expenditures for trials in the 5-year period covered. (See Section 2.1 for comments on differences in the type of trials undertaken by the two institutes.)

The total projected expenditures[1] for clinical trials are shown in Table 6-4 by FY. Results in the table are given as a percentage of the total projected expenditures for all institutes combined. The percentage distribution for FY 1979 expenditures (Table 6-3) was about the same as for FY 1979 projected expenditures (Table 6-4). This was not true for FY 1975 through FY 1978. Some of the change was due to the length of trials sponsored by the NHLBI and NCI. The average length of NCI trials listed in the 1975 Inventory was 2.47 years, contrasted with [illegible] years in the 1979 Inventory. The corresponding figures for NHLBI trials were 3.58 and [illegible], respectively.

1. Previous expenditures plus projected future expenditures for trials counted in Table 6-1.

Table 6-1 Number of NIH-sponsored trials, by institute and fiscal year

Institute	*Fiscal year (FY)* 1975	1976	1977	1978	1979
Cancer (NCI)	405	522	418	515	654
Eye (NEI)	20	21	22	28	26
Allergy and Infectious Diseases (NIAID)	109	141	93	99	120
Arthritis, Diabetes, and Digestive and Kidney Diseases (NIADDK)	49	50	49	51	67
Child Health and Human Development (NICHD)	41	52	53	39	32
Dental Research (NIDR)	44	34	36	37	26
General Medical Services (NIGMS)	2	0	0	1	1
Neurological and Communicative Diseases and Stroke (NINCDS)	59	73	51	55	40
Heart, Lung, and Blood (NHLBI)	26	26	24	20	20
All NIH	755	926*	746	845	986

*Includes 7 trials done in the NIH Clinical Center.

Table 6-5 provides total projected expenditures per patient-year of study for FY 1979 trials. This figure, for a given trial, was derived by dividing the total projected expenditures for the trial by the product of the projected sample size and the number of years the trial was expected to run. This calculation was made for each trial listed in the Inventory. The resulting values were ranked from lowest to highest. The value falling at the 50th percentile constituted the median projected expenditure per patient-year of study. The mean projected expenditure per patient-year was calculated by dividing the total projected expenditures for all trials by the sum of products derived by multiplying the projected sample size and expected duration of the individual trials.

Note that both the median and mean are underestimates of the actual per patient-year expenditures since they are derived under the assumptions that the full complement of patients, as given by the projected sample size, is enrolled as soon as the trial is funded and that it remains under follow-up to the end of funding for the study. Neither assumption is likely to be true. However, more refined calculations were not possible with the data provided.

The median expenditure per patient-year for FY 1979 trials was $574 and ranged from a low of $70 to a high of $1,657. The mean expenditure was $273 and ranged from $31 to $889. (See Chapter 4 and Meinert, 1982, for discussion of expenditures for single-center versus multicenter trials.)

Table 6-5 also provides sample size data. The median sample size of all 986 trials was 100 (range 30 to 850). The mean was 670 (range 99 to 2,589).

Table 6-6 provides expenditure data[2] from 1970 through 1981 for Veterans Administration

2. From the Veterans Administration Cooperative Studies Program, VA Central Office, Washington, D.C., 1981.

Table 6-2 NIH expenditures for clinical trials as a percentage of total NIH appropriations

	Fiscal year (FY) 1975	1976	1977	1978	1979
A Total NIH appropriations (millions $)	$2,093	$2,302	$2,544	$2,843	$3,190
B NIH expenditures* for clinical trials (millions $)	$88	$121	$105	$122	$136
C Percent of total (i.e., B ÷ A × 100)	4.2	5.3	4.1	4.3	4.3

*Includes general support provided to the Division of Research Resources of the NIH and to the NIH Clinical Center.

Table 6-3 Percent distribution of total NIH expenditures for clinical trials, by institute and fiscal year

Institute	Fiscal year (FY) 1975	1976	1977	1978	1979
Cancer (NCI)	30.2	34.7	35.9	31.9	[illegible]
Eye (NEI)	3.5	3.9	4.4	5.3	[illegible]
Allergy and Infectious Diseases (NIAID)	3.5	4.1	2.8	3.1	[illegible]
Arthritis, Diabetes, and Digestive and Kidney Diseases (NIADDK)	3.8	6.4	6.1	6.6	[illegible]
Child Health and Human Development (NICHD)	4.4	5.0	4.3	3.1	[illegible]
Dental Research (NIDR)	2.0	1.3	2.7	2.5	[illegible]
General Medical Services (NIGMS)	0.1	0.0	0.0	0.2	[illegible]
Neurological and Communicative Diseases and Stroke (NINCDS)	3.9	2.4	2.6	2.5	2.0
Heart, Lung, and Blood (NHLBI)	48.6	42.1	41.1	44.9	[illegible]
All NIH	100.0	100.0	100.0	100.0	100.0
Total NIH expenditures for clinical trials (millions $)	$87.8	$120.6*	$105.3	$122.3	[illegible]

*Includes expenditures for 7 trials done in the NIH Clinical Center.

(VA) sponsored multicenter trials. The support for such trials represented a little over 3% of the total VA research and development (R and D) budget in 1970, contrasted with slightly over 7% in 1981. The portion of VA research funds awarded to individual centers to conduct single-center trials was not available.

6.2 WHO SHOULD FINANCE CLINICAL TRIALS?

Clearly, the federal government via the NIH, VA, or other agencies can provide only a fraction of the support needed to carry out clinical trials. In fact, there is concern that the present level of government funding is already too high and that the support is siphoning funds from other more worthy areas of research.

In an ideal world, the drug and device industry would underwrite the costs for establishing both the efficacy and long-term safety of proprietary products. Government support would be limited primarily to commercial products that offer manufacturers little or no opportunity for profit. Health insurance carriers, such as Blue Cross and Blue Shield, as well as Medicare and Medicaid would support trials designed to evaluate specific health care procedures, as well as trials aimed at assessing the cost effectiveness of different methods of health care delivery.

We are still a long way from the ideal. Drugs such as the hypoglycemic agents have been marketed without any evidence of long-term safety or efficacy in relation to the prime reason for their continued use—reduction of morbidity and premature death associated with diabetes. Most of the data on the long-term safety and efficacy of proprietary drugs used for chronic conditions, such as diabetes and heart disease, have been assembled at government expense.

Health insurance carriers and their clients, instead of encouraging trials, have payment policies that discourage them. The general prohibition against payments for "experimental" procedures in most health insurance plans leads to the paradox in which coverage may be denied when a procedure is being tested as part of a clinical trial but not when that same procedure is used by practitioners outside the context of any trial.

The drug prescribing practices of the medical profession have an effect on the testing and licensing practices of the drug industry. It is clear that physicians prescribe drugs for purposes

Table 6-4 Percent distribution of total NIH projected expenditures* for clinical trials, by institute and fiscal year

Institute	Fiscal year (FY) 1975	1976	1977	1978	1979
Cancer (NCI)	20.6	23.2	22.9	24.2	[illegible]
Eye (NEI)	3.1	3.3	7.4	7.7	[illegible]
Allergy and Infectious Diseases (NIAID)	2.0	2.9	2.2	2.2	[illegible]
Arthritis, Diabetes, and Digestive and Kidney Diseases (NIADDK)	5.4	6.2	5.8	6.0	[illegible]
Child Health and Human Development (NICHD)	3.0	3.8	3.3	2.9	[illegible]
Dental Research (NIDR)	1.8	1.6	1.6	1.7	1.0
General Medical Services (NIGMS)	0.0	0.0	0.0	0.0	0.0
Neurological and Communicative Disorders and Stroke (NINCDS)	2.8	3.1	2.3	2.5	[illegible]
Heart, Lung, and Blood (NHLBI)	61.4	55.9	54.5	52.7	[illegible]
All NIH	100.0	100.0	100.0	100.0	100.0
Total projected expenditures for clinical trials (millions $)	$641.8	$739.3†	$848.6	$848.4	[illegible]

*Includes expenditures through the indicated fiscal year plus projected future expenditures (see Table 6-1 for counts)

†Includes expenditures for 7 trials done in the NIH Clinical Center.

Table 6-5 Mean and median projected expenditures* per patient-year of study for trials listed in the 1979 Inventory

Institute	Number of trials	Sample size Mean	Sample size Median	Projected expenditure per patient-year Mean	Projected expenditure per patient-year Median
Cancer (NCI)	654	269	100	$237	$ 603
Eye (NEI)	26	482	200	$706	$ 350
Allergy and Infectious Diseases (NIAID)	120	1,373	100	$ 31	$ 302
Arthritis, Diabetes, and Digestive and Kidney Diseases (NIADDK)	67	180	70	$674	$1,036
Child Health and Human Development (NICHD)	32	473	100	$383	$ 483
Dental Research (NIDR)	26	943	663	$ 55	$ 70
Neurological and Communicative Disorders and Stroke (NINCDS)	40	99	30	$889	$1,155
Heart, Lung, and Blood (NHLBI)	20	2,589	850	$873	$1,657
All NIH	986†	670	100	$273	$ 574

*Includes expenditures through FY 1979 plus projected future expenditures.

†Includes 1 trial sponsored by the National Institute of General Medical Sciences.

Table 6-6 VA expenditures for multicenter clinical trials, by fiscal year

Fiscal year	Total R and D budget*	Multicenter clinical trials*	Cost as percent of total R and D budget
1970	$ 58.1	$1.8	3.1
1971	$ 60.9	$1.8	3.0
1972	$ 69.1	$1.8	2.6
1973	$ 78.6	$2.4	3.1
1974	$ 81.8	$4.3	5.3
1975	$ 95.4	$5.4	5.7
1976†	$101.6	$5.9	5.8
1977	$109.6	$5.8	5.3
1978	$118.0	$6.3	5.3
1979	$126.3	$8.5	6.7
1980	$137.7	$9.0	6.5
1981	$137.5	$9.7	7.1

*In millions of dollars.

†Adjusted for switch in starting date for fiscal year from July 1 to October 1.

other than the approved indications (Committee on Drugs, 1978; Erickson et al., 1980; Mundy et al., 1974). The sales spurt following approval of cimetidine (Tagamet®) in 1977 for use with duodenal ulcer and Zollinger-Ellison syndrome is a case in point. The spurt was due in large measure to use of the drug for unapproved indications. A total of 2,840 patients were identified as having received cimetidine in two Baltimore area hospitals from July 1978 to January 1979 (Cocco and Cocco, 1981). Among this number, only 604 (21%) had established diagnoses for the two approved indications. A survey by Schade and Donaldson (1981) involved 200 consecutive patients admitted to the Yale University Hospital and the West Haven Veterans Administration Medical Center (100 patients from each of the two institutions) who received a prescription for cimetidine. Only 15 of the patients (7.5%) were given the drug for an approved indication. The authors concluded that:

> *Our findings strongly suggest that physicians now prescribe cimetidine for remarkably diverse purposes, most of which have not been validated.*

Why should a drug company undertake the expense of testing an established drug for a new indication if it is already being used for that indication?

The Food and Drug Administration (FDA) approval process for a drug to be used with a chronic condition, such as elevated blood glucose or lipid levels, requires the manufacturer to show only that the proposed drug is safe and effective (e.g., in the case of a hypoglycemic agent, that it lowers blood glucose levels). Evidence of effectiveness in reducing morbidity or mortality associated with the condition is not required. Others, outside the drug industry, via government funded trials such as the UGDP and CDP, have had to gather the evidence (see Coronary Drug Project Research Group, 1973a; University Group Diabetes Program Research Group, 1970d).

Even the patent law that protects proprietary drugs may serve to reduce incentives for industry-sponsored long-term trials. Protection is limited to a 17-year period. Proprietary products can be marketed by other manufacturers under their own trade names once the period of protection expires. The period for protected sales will be less, sometimes much less, than the 17 years after deducting time needed by the manufacturer to test the drug and obtain approval from the FDA for marketing the drug.

There are proposals before the United States Congress to extend the period of protection [illegible] they have not yet been acted upon. The [illegible] legislation involving so-called orphan drugs [illegible] example of the importance of the legislative process in facilitating the development of drugs [illegible] this case for rare diseases that offer little opportunity for industry profit (Finkel, 1982).

Mechanisms need to be developed that [illegible] facilitate the mixture of public and private [illegible] for conduct of worthwhile trials. Drug firms [illegible] provide limited support for some government-sponsored trials, via drugs, devices, and other materials they supply free of charge. However, they will be reluctant to provide massive financial aid unless the leadership of the study [illegible] responsive to their needs in the FDA approval process. A prototype organizational structure [illegible] required. In fact, many of the necessary organizational principles have already been developed. For example, the organizational guidelines [illegible] ensuring a separation of functions in PARIS (Persantine Aspirin Reinfarction Study Research Group, 1980a) were similar to those used in AMIS (Aspirin Myocardial Infarction Study Research Group, 1980a). The latter trial was government funded; the former was privately funded.

Private health insurance companies and their clients must be encouraged to take a more positive approach toward the support of worthwhile trials. Investments of this sort could pay dividends in reduced costs for health care insurance in the future, if coverage for new procedures were denied until or unless they were shown to be of benefit via properly designed and executed trials. The NIH, even with greatly expanded resources, cannot be expected to bear the full burden of these costs and still provide needed support for basic research. Other resources are required if the momentum developed in the 1970s for planned evaluations is to be continued into the eighties and beyond.

Expenditures for health care have increased at an average rate of nearly 12% per year during the last two decades, as contrasted with [illegible] the gross national product for the same period (Weichert, 1981). Expenditures totaled $24[illegible] billion in 1980 with $20 billion for Medicaid and $28 billion for Medicare in FY 1979 (Department of Health and Human Services, 1981). Expenditures for trials aimed at evaluation of [illegible]ated health care procedures are minuscule [illegible] comparison. There is need for a more realistic [illegible] Creation of a fund pegged at just 1% of the U.S. expenditure for health care would have [illegible] an evaluation budget of nearly $2.5 billion in 1980. Contrast that with the $136 million expenditure in FY 1979 for NIH-sponsored clinical trials (Table 6-2).

6.3 FACTORS THAT INFLUENCE THE COST OF A TRIAL

6.3.1 Design

A trial, especially when carefully designed and executed, can be a costly undertaking. The need for cost efficiency is obvious, particularly in an era of shrinking budgets and skyrocketing costs. Factors influencing cost include:

- Patient eligibility criteria
- Number of patients required for study
- Time required to develop the study protocol and data collection forms
- Outcome variable to be used to measure success of the treatments
- Number of clinics and speciality resource centers required for the trial
- Treatment procedures to be used
- Ease of patient identification and enrollment
- Complexity and frequency of data collection
- Length of follow-up
- Frequency of follow-up contacts and examinations
- Time required for final data analysis
- Time required to close out the study

The frequency of patient contacts and the amount of data collected per contact is a major cost determinant. A trial requiring treatment administration over an extended time period and an outcome measure that can be observed only at regular clinic visits will require a more elaborate follow-up examination schedule than one involving a short period of treatment and death or some other easily diagnosed event as the outcome measure. The Physicians' Health Study—PHS (Sketch 1, Appendix B) is an example of a long-term drug trial not involving any direct patient contact. Patients—in this case physicians—are recruited via mail. Those who agree to participate receive their assigned medication (daily doses of aspirin, aspirin and beta-carotene, or placebo) in the mail. Follow-up for mortality is done via the National Death Index (National Center for Health Statistics, 1981) or via patients' families.

No-contact designs, such as that used in the PHS, can be considered only under special circumstances. General conditions required include use of:

- A reliable, easily observed outcome measure
- Treatments that have few side effects or complications
- Entry criteria that are not dependent on clinical assessments
- A literate, reasonably sophisticated study population

6.3.2 Planning

Starting a trial with an ill-conceived research plan or inadequately tested data forms can result in a waste of money. Serious design mistakes may make it necessary to abort the trial. Even if such drastic action is not needed, modifications to the data collection procedures after the trial is under way can be costly to implement, especially when the formats of data that have already been collected must be changed to render them compatible with revised formats. A cost element that is often underestimated is that of data processing and analysis. Underfunding this activity can seriously hamper the entire data collection process (see Chapter 5 for a discussion of data center costs).

It is not uncommon for long-term trials to cost more than originally anticipated. This can be illustrated with trials sponsored by the NHLBI, although the problem is not unique by any means to this Institute. Among the NHLBI trials appearing in both the 1975 and 1979 NIH inventories of clinical trials, only one reported a lower projected cost in 1979 than in 1975. The projected total expenditures given in 1979 were more than double the figures given in 1975 for three of the trials. Some of the changes undoubtedly were due to failure to anticipate inflationary trends over the 5-year period. However, most of the increases were too large to be explained by inflation.

One reason for increased costs has to do with shortfalls in patient recruitment and the actions taken to make up for the shortfalls via more intensive recruitment efforts and extensions of the periods of follow-up. A paper published by investigators in the Cooperative Studies Program of the Veterans Administration reviewed the recruitment performance of seven multicenter trials supported by that program (Collins et al., 1980). One trial was terminated due to recruitment problems. None of the other six trials were able to complete recruitment within the time frame originally proposed. All six required extensions for patient recruitment or had to settle for fewer patients than originally planned. Even with extensions, none of the trials achieved the original sample size goal.

6.3.3 Multipurpose studies

It is not unusual for a trial to be designed to satisfy a number of secondary objectives in addition to the primary one. A common one relates to the description of the natural history of the disease under treatment in long-term trials, such as the CDP (Coronary Drug Project Research Group, 1973a). The addition of secondary objectives can add to the cost of the trial. The increase will be smallest for objectives that can be pursued with data needed for the primary objective as well, and largest when added data are needed. The decision as to whether to pursue secondary objectives should depend on the scientific importance of those objectives, the suitability of the trial as a vehicle for pursuing them, the chances of successfully achieving them, and the costs associated with their pursuit.

6.3.4 Ancillary studies

The trial, especially in a large multicenter trial, may provide investigators with opportunities for a number of ancillary studies (see Glossary for definition). Some may involve added patients, whereas others may simply require special analyses of existing data. However, as with pursuit of secondary objectives, they can add to the cost and complexity of the trial. Priorities should be given to those studies that are needed to understand the action of the treatments under study and to those concerning methodological issues of direct importance to the trial. No study should be undertaken that jeopardizes pursuit of the primary objective.

6.3.5 Equating the data collection needs of the trial with those for patient care

The data required to satisfy the research aims of the trial may be different from those needed for patient care. Failure to distinguish data needed for this latter purpose from those needed for the trial can lead to the collection of superfluous information that is a burden to collect and process.

6.3.6 Undisciplined data collection philosophy

The data collection schedule for the trial should be kept as simple as possible. Strong leadership is required to ensure the development of a focused data collection philosophy and related set of data forms. Without this leadership, the data collection scheme can be a hodgepodge of peripherally related data items designed to cater to the special interests of specific investigators in the trial.

6.4 COST CONTROL PROCEDURES

6.4.1 General cost control procedures

Cost control is the combined responsibility of the sponsor and study investigators. There is no substitute for a cost-conscious investigator group. Some of the more obvious extravagances to be avoided are:

- Use of costly state-of-the-art technology when less sophisticated technology will suffice
- Unnecessary travel at study expense
- Use of study funds for lavish office furnishings or for activities not related to the trial
- Overstaffing

"Cost saving" measures to be avoided include:

- Submission of an unrealistically low budget request in the hope of improving the prospects for funding
- Undue reliance on existing staff paid from other sources to perform essential functions in the trial
- Cutbacks on financial support for data analysis in order to increase support for data collection activities
- Reduction of the sample size requirement for the trial by switching from a single event to a composite of events or to a laboratory measure as the outcome measure
- Changing the sample size calculation so as to bring it in line with the number of patients available for study
- Sponsor-imposed travel restrictions in a multicenter trial that limit the ability of investigators to interact and function as a cohesive unit

6.4.2 Method of funding

The funding structure for the trial will in itself provide some cost controls. Ceilings placed on expenditures when awards are made, as with most NIH grant awards, encourage the conservation of funds, provided unused funds accrued in one year can be carried over for use in the next year. Awards with cost-reimbursement features, as with some NIH contracts, generally include provisions for periodic cost reviews by the sponsor over the life of the award (see Chapter 21 for additional discussion).

The differences between grant and contract methods of funding are most apparent in the budgeting process. An investigator is required to submit a budget for the specified number of years before the start of the trial with a fixed cost multiyear grant. Budgeting is done with the realization that the funds requested may be reduced if the budget is perceived as excessive by reviewers of the proposal. Approved applications that are funded are supported up to, but not above, the approved ceiling figures set when awards are made. An investigator who has done a poor job of anticipating costs for the trial will have to cut back on activities planned or seek supplemental funds to make up for deficits.

The budget preparation process is different for cost-reimbursement contracts. Costs can exceed the original budget and still be recovered. However, reliance on the cost-reimbursement mode of funding can pose dilemmas for investigators when preparing their initial budget requests in conjunction with Request for Proposals (RFP). Submission of a realistic budget that includes support for activities deemed necessary by the investigator but not mentioned in the RFP may cause the response to be viewed as noncompetitive. Realization of this fact may tempt him to adopt a more "pragmatic" approach to the budgeting process (i.e., by preparing a budget which he believes to be in the competitive range, even if he considers it to be too small), since the costs for "unanticipated" but justifiable activities can be recovered later as part of the cost-reimbursement process.

Funding is tied to the actual level of activities in the cost-reimbursement approach. This is more difficult to do with fixed-cost awards. One method of funding that combines features of the two approaches, at least for clinics, involves awarding a designated amount for fixed costs, plus a variable sum that depends on numbers of patients enrolled and followed. However, a word of warning is in order. Capitation forms of payment can lead to questionable practices if clinic personnel are tempted to cut corners in order to ensure an adequate flow of patients to maintain a desired level of funding.

6.4.3 Cost reviews

The investigator cannot develop or maintain a cost-conscious attitude without periodic reviews of activities and their associated costs. Such reviews are especially important in trials involving two or more primary work components, such as in CASS (Coronary Artery Surgery Study Research Group, 1981). That study required a separation of the coordinating center costs for the trial and registry components of the study. The separation was used as a management tool to make certain that data intake and analysis priorities were met for both components.

6.4.4 Periodic priority assessments

The usual approach is to add new data collection and quality control procedures as they are needed over the course of the trial, without much thought regarding their importance in meeting the main objectives of the trial (Meinert, 1977). Periodic revisions and prunings performed by the leadership of the trial are necessary if the procedures are to remain lean and efficient.

6.4.5 Review and funding for ancillary studies

The study leadership should develop an internal review process for proposed ancillary studies (see Glossary). Only those studies that do not interfere with patient recruitment, data collection, or other essential activities in the trial, should be approved. Studies that are too costly to undertake without additional funding should be reviewed subject to acquisition of funding.

Ancillary studies, by definition, are designed to address questions that are of secondary or peripheral importance to the main objectives of the trial. However, since they are done by investigators involved in the trial and are often carried out on subgroups of study patients, they can add to both the cost and the complexity of the trial. They may even compromise the ability of the investigators to pursue the main aims of the trial. Part of the purpose of the review process is to make certain that this does not happen and to ensure that the investigations do not siphon away resources needed for the trial itself. Small amounts of support, particularly in the form of study staff, may be derived from the trial. Undertakings requiring added staff should be funded and operated independently of the trial.

6.4.6 Justification of data items

The data collection requirements of the trial should be limited to those that are directly related to the aims of the trial and should not be confused with other needs, such as those required for patient care or for ancillary studies. Every item that appears on the data forms should be required for pursuit of one of the aims of the trial. Items that cannot be justified in this manner should not be made part of the official data set of the trial.

6.4.7 Use of low-technology procedures

The cost of a trial will be influenced by the level of technology needed for the procedure used in the trial. Insistence on high-technology procedures can result in a significant increase in expenses, especially if special equipment must be purchased and skilled personnel hired to operate it. State-of-the-art instrumentation is generally not essential to the success of most trials.

6.5 NEED FOR BETTER COST DATA

Reliable data on the costs of trials are difficult to obtain. Expenditure records maintained by the NIH are too crude to permit anything more than a rough analysis of cost (Meinert, 1979a). Cost comparisons across governmental agencies, such as the NIH and VA, are further complicated by differences in funding and accounting practices. For example, NIH-sponsored trials typically include salary support for senior as well as essential support staff, whereas personnel costs in VA-sponsored trials are generally limited to those needed for essential support staff. Comparisons between countries are even more difficult to make. For example, studies done in the United Kingdom always appear to be less expensive than in the United States because of fundamental differences in the way health care procedures are paid for in the two countries.

Reliable cost data for industry-sponsored trials are even more difficult to obtain. A for-profit business firm is not eager to provide detailed research expenditure data for review by the general public or competing firms.

Nevertheless, designers of trials need to have a better understanding of the way in which costs accumulate and how they are influenced by factors under the designers' control, especially in relation to the types and amounts of data collected. This understanding can only be achieved through the collection of detailed cost data related to specific data collection and analysis activities in a variety of trials.

7. Impact of clinical trials on the practice of medicine

A new scientific truth does not triumph by convincing its opponents and making them see the light, but rather because its opponents eventually die, and a new generation grows up that is familiar with it.

Max Planck

7.1 INTRODUCTION

There is need for a better understanding of the way trials influence the practice of medicine. What is their role in establishing new treatments or in discrediting old ones? When can they be expected to play a role and when not? Does the design or the way in which a trial is executed influence the way it is perceived—in the medical community and by the lay public? Answers to questions of this kind could promote the design of better, more potent, trials in the future. (See references 59 and 366 for additional discussion.)

7.2 FACTORS INFLUENCING TREATMENT ACCEPTANCE

7.2.1 Prior opinion and previous experience with a treatment

A treatment that has been around for a long time, even if trials have shown it to be of no value, will fade from favor more slowly than one still in its infancy. Chalmers has noted the continued use of bed rest in the treatment of acute viral hepatitis after several trials, all of which have failed to indicate any merit for the treatment. Similarly, ulcer patients continue to be placed on "sippy" diets, even though trials have failed to show the value of such diets (Chalmers, 1974).

The time to do a trial is before the treatment is accepted as standard practice. It will be difficult to mount one once that has happened. For example, it would be quite difficult to mount trials now to evaluate the efficacy of coronary care units (CCU) in the treatment of acute myocardial infarction (MI) victims. The units are presumed to be of value. Assigning patients to a CCU or regular hospital care at random might well be regarded as a questionable practice in today's climate.

7.2.2 Clinical relevance of the outcome measure

All other things being equal, a trial with death or some other serious morbid event as the outcome should receive more attention than one involving less relevant outcomes. It is distressing, in this context, to note the number of trials that rely on nonclinical measures, such as a laboratory test, to evaluate a treatment (see Chapter 2).

7.2.3 Degree to which test treatment simulates real-world treatment

Ideally, the test treatment should be used in the exact same manner as in the real world. However, this is not always possible. The need for uniformity in the treatment process makes it necessary to impose conditions on usage not ordinarily encountered in real life. For example, drugs may have to be given in a single fixed dose in double-masked trials, even though they are not used this way in practice.

7.2.4 Consistency of findings with previous results

The judgment regarding the virtues of a treatment should be based on a digest of all pertinent data—not only the last report. Survey papers, such as those produced by Chalmers and coworkers (1972, 1977), represent examples of efforts aimed at amalgamating information from several trials to assess the merits of a treatment.

It is desirable to have several replications of a trial before reaching a conclusion regarding a treatment. Unfortunately, the world is usually not so obliging. The high cost of some trials, such as the Multiple Risk Factor Intervention Trial (in excess of $100 million), makes it impractical to consider replication. Replication in other cases may be ruled out on ethical grounds. For example, it would be impossible to replicate the Veterans Administration (VA) studies on frank hypertensives. No physician would be prepared to have such patients assigned to a placebo treatment (Veterans Administration Cooperative Study Group on Antihypertensive Agents, 1967, 1970).

7.2.5 Direction of results

The direction of the trial results will influence the way in which they are received. It is easier to accept a positive finding than a negative one, especially if the finding pertains to an "established" treatment. Physicians are trained to be more comfortable giving a treatment than withholding one. Patients as well usually find it more consoling to receive a treatment than to be denied one.

7.2.6 Importance of the treatment

The interest generated by a particular trial will be influenced by the number of persons in the medical community who regard the treatment as useful. The attention accorded the UGDP findings was much greater than that for the Coronary Drug Project (CDP). Undoubtedly, the difference was due in part to the fact that the treatments used in the UGDP were established modes of therapy for the mild, noninsulin-dependent diabetic, whereas this was not the case for the drugs used in the CDP for patients with a prior MI.

7.2.7 Cost and payment schedule

The cost of the treatment and the opportunity for covering those costs from third-party sources, such as insurance carriers, will play a role in treatment "acceptance." Use of dialysis for end-stage renal disease is a case in point. The big spurt in use of the treatment came with enactment of legislation in 1972 that provided payment for the procedure from Social Security funds. The number of people on dialysis in the United States jumped from 2,400 in 1970 to nearly 27,000 by 1977 and to over 44,000 by 1979 (Burton and Hirschman, 1979a, 1979b).

7.2.8 Treatment facilities and resources

The opportunities for administering a treatment will be limited by the nature of staff and support facilities needed for its administration. Heart transplantation is a case in point. The utility of the treatment is limited by organ availability, the number of trained transplant teams, and available support facilities.

7.2.9 Design and operating features of the trial

Ideally, the weight given to a result should be determined by an unbiased, objective evaluation of the strengths and weaknesses of the trial. In fact the evaluation may be done carelessly and from a preconceived point of view. Design or operating features regarded as major weaknesses of one trial may be overlooked or ignored in another, depending on the direction of the results. Evidence of such double standards can be seen from a comparison of the criticisms directed at the UGDP study of tolbutamide with those directed at studies done by Keen and by Paasikivi (Keen and Jarrett, 1970; Keen, 1971; Paasikivi, 1970). The UGDP results were negative, whereas the other two were considered to be positive.

7.2.10 Study population

The degree to which the study population approximates a real-life mix of patients may influence the way results are received. A clinician's perception that patients treated in the trial were radically different from those he treats may lead him to downplay or completely reject the results.

7.2.11 Method of presentation

Treatment acceptance can be influenced by the way in which results are presented. Negative attitudes that develop in the medical community because of the mode of presentation may cause its members to reject the findings for emotional reasons. There is some evidence that this happened with the UGDP. The tolbutamide findings were presented at a national meeting of the American Diabetes Association (ADA) in June 1970. The paper containing the results first appeared 5 months after the presentation, and then only in a speciality journal with limited circulation (University Group Diabetes Program Research Group, 1970e). The press coverage following the presentation resulted in a deluge of inquiries to practicing diabetologists around the country regarding the treatment. Many of them resented having to deal with the questions before the results were published.

The potential for ill will is not limited to trials with negative findings, as may be seen in the Macular Photocoagulation Study (MPS) with presentation of results for treatment of senile macular degeneration (Macular Photocoagulation Study Group, 1982, 1984). The study avoided the UGDP publication lag by mailing a preprint of the manuscript to all practicing ophthalmologists in the U.S. The National Eye Institute scheduled a press conference a few days after the mailing and just before the manuscript appeared in print. The national TV coverage of the results took many treating ophthalmologists by surprise, particularly those who had not yet received the paper or who had not read it. The public relations problem might have been avoided if there had not been a press conference, but public awareness of the results was considered to be essential because of the need for patients to recognize the symptoms of senile macular degeneration so as to obtain early diagnosis and treatment.

7.2.12 Counterforces

There may be a number of counterforces working against the acceptance of a finding. Such forces can be expected to emerge whenever results run contrary to established dogma, and especially when major financial considerations are involved. A medical specialist whose practice depends on the treatment being questioned will be much more reluctant to accept negative findings than positive findings. The Committee for he Care of the Diabetic was formed by a group of diabetologists largely as a means of counteracting the UGDP findings and the proposed Food and Drug Administration (FDA) labeling changes for the oral hypoglycemic agents (see Section 7.4).

The drug company whose product is threatened by the study can be expected to question the findings and to express doubts regarding the study. These expressions may take the form of prepared press releases indicating that the trial should not be regarded as definitive and making the universal call for further research. Upjohn, the manufacturer of Orinase® (tolbutamide), as well as other manufacturers of hypoglycemic agents, sent "Dear Doctor" letters to practicing diabetologists warning of the need for caution when interpreting the findings of the UGDP (see Knox, 1971, and Mintz, 1970b for references to the letters). Consultants were hired by Upjohn to critique the study and to speak at meetings where the findings were discussed. Company sales personnel were provided with "informational material" for answering questions concerning the study. The material summarized criticisms of the study and reminded physicians of other work supportive of the treatment.

Another force with interests allied to the pharmaceutical firms is that associated with the so-called "throw-away" medical journals.[1] Such publications rely heavily on advertising from drug manufacturers for their income (Chalmers, 1982a; Warner et al., 1978). The editorial policy of publications such as the *Medical Tribune* and the *Hospital Tribune* was negative, if not downright hostile, toward the UGDP, while carrying ads for hypoglycemic agents.

7.3 IMPACT ASSESSMENT

Changes in health care practices occur gradually and for a variety of reasons. Methods used to relate such changes to specific events, such as the publication of results from a particular trial, are at best approximate. It is always dangerous to associate any change involving complex behaviors with any single event. A case in point is the growing emphasis on the diagnosis and treatment of hypertension. Unquestionably, the emphasis stems, at least in part, from trials supporting the value of antihypertensive treatment. But it is also due to massive efforts by the federal government and the medical profession to alert the public to the dangers of hypertension. Communities throughout the nation have carried out screening programs to identify hypertensives. The National High Blood Pressure Education Program, founded in 1972 and sponsored by the National Heart, Lung, and Blood Institute (NHLBI), has been aimed at educating members of the public and the medical community to the importance of blood pressure control (National Heart, Lung, and Blood Institute, 1973; Szklo, 1980). Physician visits during which at least one antihypertensive drug was prescribed increased by about a third from 1968 to 1978 (from data provided in the National Disease and Therapeutic Index, IMS America Ltd.,[2] Ambler, Pennsylvania). There was a 27% decrease in mortality rates for coronary heart disease over the same time interval (Working Group on Arteriosclerosis, 1981).

In the light of such evidence, it is tempting to attribute the decline to more aggressive treatment resulting from trials and educational programs. However, those who do so ignore the fact that mortality due to cardiovascular disease was already on the decline before the first VA hypertension trials started and before widespread public awareness of the dangers of hypertension.

Prescription and sales data can be used to provide gross indications of changes in treatment patterns. Data from IMS are used in Section 7.4 to chart changes in the use of oral hypoglycemic agents from 1964 forward.

Other indications of change may be obtained from other data sources, such as the Professional Services Review Organization (PSRO) or from the Commission on Professional and Hospital Activities (CPHA). The Commission is based in Ann Arbor, Michigan, and maintains a variety of usage statistics for member hospitals. Payment data maintained by private health insurance carriers and by Medicare and Medicaid also can be helpful in tracing treatment patterns.

More direct measures of change can be obtained from special surveys, such as the one done by Stross and Harlan (1979) designed to assess the awareness of primary-care physicians regarding results from the Diabetic Retinopathy Study—DRS (done about 18 months after the DRS results were published). Only 28% (3[illegible] out of 137) of the family physicians and 46% of the internists surveyed (42 out of 91) were aware of the results. A similar approach was used to assess the level of physician familiarity with results from the Hypertension Detection and Follow-Up Program—HDFP (Stross and Harlan, 1981). Survey techniques also were used in a contract issued by the NHLBI to assess physician knowledge of findings from the CDP and Aspirin Myocardial Infarction Study (AMIS) (Market Facts, Inc., 1982).

7.4 THE UNIVERSITY GROUP DIABETES PROGRAM: A CASE STUDY

The UGDP was started in 1960, enrolled its first patient in 1961, completed data collection in 1975, and published its final report in 19[illegible]. Citations 464 through 470, 472, 473, 475, and 4[illegible] (Appendix I) refer to a series of original publications that detail the design, methods, and results of the study. Citations 83, 95, 161, 173, 18[illegible], 1[illegible], 192–194, 261, 386, 409, 413, 419, 459, 460, and 471, relate to the controversy that developed starting in mid-1970 with a UGDP data presentation that questioned the value of tolbutamide for use in diabetics. Table 7–1 provides a chro-

1. So termed because they are distributed to practicing physicians free of charge.

2. IMS is a private firm that specializes in the compilation of drug utilization data for sale to various business firms and agencies.

Table 7-1 Chronology of events associated with the UGDP

Year	*Month, day*	*Event*
1959	June	First planning meeting of UGDP investigators (467)*
1960	September	Initiation of grant support for the coordinating center and first 7 clinics (467)
1961	February	Enrollment of first patient (467)
1962	September	Addition of phenformin to the study and recruitment of 5 additional clinics (467)
1966	February	Completion of patient recruitment (467, 468)
1969	June 6	UGDP investigators vote to discontinue tolbutamide treatment (468 and UGDP meeting minutes)
1970	May 20	Tolbutamide results on Dow Jones ticker tape (327)
1970	May 21, 22	*Wall Street Journal, Washington Post,* and *New York Times* articles on tolbutamide results (280, 326, 408)
1970	June 14	Tolbutamide results presented at American Diabetes Association meeting, St. Louis (464, 465, 466)
1970	October	Food and Drug Administration (FDA) distributes bulletin supporting findings (179)
1970	November	Tolbutamide results published (468)
1970	November	Committee for the Care of Diabetics (CCD) formed (183)†
1971	April	Feinstein criticism of UGDP published (161)
1971	May 16	UGDP investigators vote to discontinue phenformin treatment in UGDP (470, 472, and UGDP meeting minutes)
1971	June	FDA outlines labeling changes for sulfonylureas (180)
1971	August 9	UGDP preliminary report on phenformin published (470)
1971	September 14	Associate Director of National Institutes of Health (NIH) asks president of International Biometrics Society to appoint a committee to review UGDP (83)
1971	September 20	Schor criticism of UGDP published (409)
1971	September 20	Cornfield defense of UGDP published (95)
1971	October 7	CCD petitions commissioner of the FDA to rescind proposed label change (183 and actual petition)
1972	May	FDA reaffirms position on proposed labeling change (181)
1972	June 5	FDA commissioner denies October 1971 request to rescind proposed label change (183)
1972	July 13	CCD requests evidentiary hearing before FDA commissioner on proposed labeling changes (183)
1972	August 3	Commissioner of FDA denies CCD request for evidentiary hearing (451)
1972	August 11	CCD argues to have the FDA enjoined from implementing labeling change before the United States District Court for the District of Massachusetts (451)
1972	August 30	Request to have the FDA enjoined from making labeling change denied by Judge Campbell of the United States District Court for the District of Massachusetts (183, 451)
1972	August	Biometrics Society Committee starts review of UGDP and other related studies (83)
1972	September	Seltzer criticism of UGDP published (419)
1972	October 17	Second motion for injunction against label change filed by CCD in the United States District Court for the District of Massachusetts (451)
1972	October	Response to Seltzer critique published (471)
1972	November 3	Temporary injunction order granted by Judge Murray of the United States District Court for the District of Massachusetts (451)
1972	November 7	Preliminary injunction against proposed label change granted by United States District Court for the District of Massachusetts (183)

Table 7-1 Chronology of events associated with the UGDP (*continued*)

Year	*Month, day*	*Event*
1973	July 31	Preliminary injunction vacated by Judge Coffin of United States Court of Appeals for the First Circuit. Case sent back to FDA for further deliberations (183, 451)
1973	October	FDA hearing on labeling of oral agents (183)
1974	February	FDA circulates proposed labeling revision (183)
1974	March–April	FDA holds meeting on proposed label change, then postpones action on change until report of Biometrics Committee (183)
1974	September 18, 19, 20	Testimony taken concerning use of oral hypoglycemic agents before the United States Senate Select Committee on Small Business, Monopoly Subcommittee (459)
1975	January 31	Added testimony concerning use of oral hypoglycemic agents before the United States Senate Select Committee on Small Business, Monopoly Subcommittee (460)
1975	February 10	Report of the Biometrics Committee published (83)
1975	February	UGDP final report on phenformin published (472)
1975	July 9, 10	Added testimony concerning use of oral hypoglycemic agents before the United States Senate Select Committee on Small Business, Monopoly Subcommittee (460)
1975	August	Termination of patient follow-up in UGDP (476)
1975	September 30	CCD files suit against David Mathews, Secretary of Health, Education and Welfare, et al., for access to UGDP raw data under the Freedom of Information Act (FOIA) in the United States District Court for the District of Columbia (452)
1975	October 14	Ciba-Geigy files suit against David Mathews, Secretary of Health, Education and Welfare, et al., for access to UGDP raw data under the FOIA in the United States District Court for the Southern District of New York (457)
1975	December	FDA announces intent to audit UGDP results (461)
1976	February 5	United States District Court for the District of Columbia rules UGDP raw data not subject to FOIA (453)
1976	February 25	CCD files appeal of February 5 decision in United States Court of Appeals for the District of Columbia Circuit (461)
1976	September	FDA audit of UGDP begins
1976	October	FDA Endocrinology and Metabolism Advisory Committee recommends removal of phenformin from market (184)
1977	March 8	United States District Court for the Southern District of New York rejects Ciba-Geigy request for UGDP raw data (458)
1977	April 22	Health Research Group (HRG) of Washington, D.C., petitions Secretary of HEW to suspend phenformin from market under imminent hazard provision of law (185)
1977	May 6	FDA begins formal proceedings to remove phenformin from market (185)
1977	May 13	FDA holds public hearing on petition of HRG (185)
1977	July 25	Secretary of HEW announces decision to suspend New Drug Applications (NDAs) for phenformin in 90 days (185)
1977	August	CCD requests that United States District Court for the District of Columbia issue an injunction against HEW order to suspend NDAs for phenformin†
1977	October 21	CCD request to United States District Court for the District of Columbia for injunction against HEW order to suspend NDAs for phenformin denied†
1977	October 23	NDAs for phenformin suspended by Secretary of HEW under imminent hazard provision of law (187)
1977	December	UGDP announces release of data listings for individual patients (474)

Table 7-1 Chronology of events associated with the UGDP (*continued*)

Year	*Month, day*	*Event*
1978	January	Appeal of October 21, 1977, court ruling filed by CCD in United States Court of Appeals for the District of Columbia Circuit
1978	July 7	Preliminary report on insulin findings published (474)
1978	July 11	Judges Leventhal and MacKinnon of the United States Court of Appeals for the District of Columbia Circuit rule that public does not have right to UGDP raw data under the FOIA. Judge Bazelon dissents (450, 461)
1978	July 25	CDC petitions United States Court of Appeal for the District of Columbia Circuit for rehearing on July 11 ruling (461)
1978	October 17	Petition for rehearing at the United States Court of Appeals for the District of Columbia Circuit denied (461)
1978	November 14	Results of FDA audit of UGDP announced (188)
1978	November 15	Commissioner of FDA orders phenformin withdrawn from market (462)
1979	January 15	CCD petitions the United States Supreme Court for writ of certiorari to the United States Court of Appeals for the District of Columbia Circuit (461)
1979	April 10	Appeal of October 21, 1977, ruling denied†
1979	May 14	Writ of certiorari granted
1979	October 31	UGDP case of Forsham et al., versus Harris et al., argued before the United States Supreme Court (462)
1980	March 3	United States Supreme Court holds that HEW need not produce UGDP raw data in 6 to 2 decision (462)
1982	April	Expiration of NIH grant support for UGDP
1982	November	Final report on insulin results published (476)
1982	November	UGDP deposits patient listings plus other information at the National Technical Information Service for public access (476, 477, 478)
1984	March 16	Revised label for sulfonylurea class of drugs released (192, 193, 194)

*Numbers in parentheses refer to citations in the Combined Bibliography (Appendix I).

†Personal communications with Robert F. Bradley, Joslin Diabetes Center, Boston, who was the first chairman of the CCD.

logy of UGDP related events (see also Appendix B for a sketch of the UGDP).

Table 7-2 provides a listing of the main criticisms of the study as offered by others and comments on their validity by the author (one of the investigators in the trial). Most of the attention was focused on the tolbutamide results because they were the first released and because of the popularity of the drug. Table 7-2 reflects this bias.

The news media carried a number of articles on the tolbutamide results, beginning with a report on May 20 appearing on the Dow Jones tickertape.[3] One article in particular, suggesting that the drug caused as many as 8,000 deaths per year,[4] created a good deal of patient anxiety and physician hostility toward the study even before the results were presented in June. (Incidentally, the number had escalated from 10,000 to 15,000 without benefit of any new data in news reports a few years later, e.g., as in the *Philadelphia Inquirer*, January 28, 1975.)

The controversy and resulting doubts about the study led to two independent audits of it. The first was undertaken by a blue-ribbon committee appointed by the International Biometrics Society and was published in 1975 (see citation 83). The second was carried out by the FDA and appeared in November, 1978 (see citation 188). Neither audit found any basis to reject the conclusions of the study.

3. The report was prepared from information in an abstract of a paper submitted to the American Diabetes Association (ADA) for presentation at its June meeting. Study investigators were surprised by the publicity before the meeting. They were not aware that it was the practice of the ADA to make the program, and the abstracts contained therein, available to the press in advance of its annual meeting.

4. The article appeared in *The Washington Post* on May 22, 1970, and in several other papers around the country over the next several days.

Table 7-2 Criticisms of the UGDP and comments pertaining to them

Criticism	*Comment*
• The study was not designed to detect differences in mortality (Schor, 1971).	• The main aim of the trial was to detect differences in nonfatal vascular complications of diabetes (UGDP Research Group, 1970d). However, this focus in no way precludes comparisons for mortality differences. In fact, it is not possible to interpret results for nonfatal events in the absence of data on fatal events.
• The observed mortality difference was small and not statistically significant (Feinstein, 1971; Kilo et al., 1980).	• It is unethical to continue a trial, especially one involving an elective treatment, to produce unequivocal evidence of harm.
• The baseline differences in the composition of the study groups are large enough to account for excess mortality in the tolbutamide treatment group (Feinstein, 1971; Kilo et al., 1980; Schor, 1971; Seltzer, 1972).	• The tolbutamide-placebo mortality difference remains after adjustment for important baseline characteristics (Cornfield, 1971).
• The tolbutamide-treated group had a higher concentration of baseline cardiovascular risk factors than any of the other treatment groups (Feinstein, 1971; Kilo et al., 1980; Schor, 1971; Seltzer, 1972).	• Differences in the distribution of baseline characteristics, including CV risk factors, is within the range of chance. Further, the mortality excess is as great for the subgroup of patients who were free of CV risk factors as for those who were not. Finally, simultaneous adjustment for major CV baseline risk factors did not eliminate the excess (UGDP Research Group, 1970e; Cornfield, 1971).
• The treatment groups included patients who did not meet study eligibility criteria (Feinstein, 1971; Schor, 1971).	• Correct. However, the number of such cases was small and not differential by treatment group. Further, analyses in which ineligible patients were removed did not effect the tolbutamide-placebo mortality difference (UGDP Research Group, 1970e).
• Data from patients who received little or none of the assigned study medication should have been removed from analysis (Kilo et al., 1980; Seltzer, 1972).	• The initial analysis included all patients to avoid the introduction of selection biases. This analysis approach tends to underestimate the true effect. Analyses in which noncompliant patients were not counted enhanced, rather than diminished the mortality difference (UGDP Research Group, 1970d).
• The data analysis should have been restricted to patients with good blood glucose control (Kilo et al., 1980).	• The analysis philosophy for this variable was the same as for drug compliance. The removal of patients using a variable influenced by treatment has a good chance of rendering the treatment groups noncomparable with regard to important baseline characteristics. In any case, analyses by level of blood glucose control did not account for the mortality difference (UGDP Research Group, 1971a).
• The study failed to collect relevant clinical data (Feinstein, 1971; Seltzer, 1972).	• The criticism is unjustified. The study collected data on a number of variables needed for assessing the occurrence of various kinds of peripheral vascular events. It is always possible to identify some variable that should have been observed with the perspective of hindsight. The criticism lacks credibility, in general and especially in this case, because of the nature of the result observed. It is hard to envision other clinical observations that would offset mortality, an outcome difficult to reverse.
• There were changes in the ECG coding procedures midway in the course of the study (Schor, 1971; Seltzer, 1972).	• Correct. However, the changes were made before investigators had noted any real difference in mortality and were, in any case, made without regard to observed treatment results (Cornfield, 1971).
• The patients did not receive enough medication for effective control of blood glucose levels (Seltzer, 1972).	• A higher percent of tolbutamide-treated patients had blood glucose values in the range indicative of good control than did the placebo-treated patients. The percentage of patients judged to have fair or good control, based on blood glucose determinations done over the course of the study, was 74 in the tolbutamide-treated group versus 59 in the placebo-treated group (UGDP Research Group, 1971a, 1976).

Table 7-2 Criticisms of the UGDP and comments pertaining to them (*continued*)

Criticism	*Comment*
• The excess mortality can be accounted for by differences in the smoking behavior of the treatment group (source unknown).	• The argument is not plausible. While it is true that the study did not collect baseline smoking histories, there is no reason to believe the distribution of this characteristic would be so skewed so as to account for the excess (Cornfield, 1971). The study did in fact make an effort to rectify this oversight around 1972 with the collection of retrospective smoking histories. There were no major differences among the treatment groups with regard to smoking. However, the results were never published because of obvious questions involved in constructing baseline smoking histories long after patients were enrolled and then with the use of surrogate respondents for deceased patients. The oversight is understandable in view of the time the trial was designed. Cigarette smoking, while recognized at that time as a risk factor for cancer, was not widely recognized as a risk factor for coronary heart disease.
• The observed mortality difference can be accounted for by differences in the composition of the treatment group for unobserved baseline characteristics (Feinstein, 1971; Schor, 1971).	• This criticism can be raised for any trial. However, it lacks validity since there is no reason to assume treatment groups in a randomized trial are any less comparable for unobserved characteristics than for observed characteristics. And even if differences do exist, they will not have any effect on observed treatment differences unless the variables in question are important predictors of outcome.
• The majority of deaths were concentrated in a few clinics (Feinstein, 1971; Seltzer, 1972).	• Differences in the number of deaths by clinic are to be expected in any multicenter trial. However, they are irrelevant to comparisons by treatment groups in the UGDP, since the number of patients assigned to treatment groups was balanced by clinic (UGDP Research Group, 1970d, 1970e).
• The study included patients who did not meet the "usual" criteria for diabetes (Seltzer, 1972).	• There are a variety of criteria used for diagnosing diabetes, all of which are based, in part or totally, on the glucose tolerance test. The sum of the fasting one, two, and three hour glucose tolerance test values used in the UGDP represented an attempt to make efficient use of all the information provided by the test (UGDP Research Group, 1970d).
• The patients received a fixed dose of tolbutamide. The usual practice is to vary dosage, depending on need (Feinstein, 1971; Schor, 1971; Seltzer, 1972).	• Most patients in the real world receive the dosage used in the study (UGDP Research Group, 1972).
• The randomization schedules were not followed (Schor, 1971).	• The Biometrics Committee reviewed the randomization procedure and found no evidence of any breakdown in the assignment process (Committee for the Assessment of Biometric Aspects of Controlled Trials of Hypoglycemic Agents, 1975).
• There were "numerous" coding errors made at the coordinating center in transcription of data into computer readable formats (Feinstein, 1971).	• There is no evidence of any problem in this regard. The few errors noted in audits performed by the Biometrics Committee and FDA audit team were of no consequence in the findings of the trial (Committee for Assessment of Biometric Aspects of Controlled Trials of Hypoglycemic Agents, 1975, Food and Drug Administration, 1978).

Table 7-2 Criticisms of the UGDP and comments pertaining to them (*continued*)

Criticism	*Comment*
• There were coding and classification discrepancies in the assembled data (Kolata, 1979).	• The coding and classification error rate was in fact low and the errors that did occur were not differential by treatment group. There were no errors in the classification of patients by treatment assignment or by vital status. Hence, the argument does not provide a valid explanation of the mortality differences observed (Committee for the Assessment of Biometric Aspects of Controlled Trials of Hypoglycemic Agents, 1975; Food and Drug Administration, 1978; Prout et al., 1979)
• The cause of death information was not accurate (Feinstein, 1971; Schor, 1971; Seltzer, 1972).	• Independent review of individual death records by the FDA audit team revealed only three classification discrepancies, only one of which affected the tolbutamide-placebo comparison (Food and Drug Administration, 1978). However, in any case the main analyses in the study and the conclusions drawn from them relate to overall mortality
• The study does not prove tolbutamide is harmful (Feinstein, 1971; Schor, 1971; Seltzer, 1972).	• Correct. It would be unethical to continue a trial to establish the toxicity of an elective treatment. Toxicity is not needed to terminate an elective treatment (UGDP Research Group, 1970d)

The FDA started work on a revised label insert for tolbutamide shortly after the results were presented in 1970. The revised label warned of potential cardiovascular complications associated with prolonged use of the drug (Food and Drug Administration, 1972a). Doubts regarding the validity of the study and concerns regarding the implications of the proposed label change led to the formation of the Committee for the Care of the Diabetic (CCD). The committee was made up of practicing diabetologists from around the country (first headed by Robert F. Bradley of the Joslin Clinic and subsequently by Peter H. Forsham of the University of California). This committee, with legal counsel, obtained a court order on November 7, 1972 staying the use of the revised label[5] (Food and Drug Administration, 1975).

A side issue of importance to the field of clinical trials—and other research fields as well for that matter—had to do with public access to UGDP raw data. Records generated by the study and housed at the UGDP Coordinating Center in Baltimore were requested on behalf of the CCD under the Freedom of Information Act—FOIA (Morris et al., 1981; Stalloney, 1982; Watson, 1981; see also Chapter 24). The request was denied by the United States District Court for the District of Columbia on February 5, 1976 (see citation 453). The decision was ultimately upheld by the United States Supreme Court in a six-to-two decision issued March 3, 1980 (see citation 462).

In spite of the controversy—or more likely because of it—the study appears to have had an effect on the treatment practices of diabetologists. It has caused both friends and foes of the study alike to re-examine the underlying rationale for treatment of the noninsulin-dependent diabetic and to consider dietary rather than pharmacological treatment of such patients (Reeves et al., 1979; West, 1980).

Sales data compiled by IMS from the National Prescription Audit[6] show a drop in the use of the oral hypoglycemic agents beginning with 1974. The estimated total number of prescriptions (new as well as refills) for all hypoglycemic oral agents in the United States has declined from a high of 21 million in 1973 to 13.6 million in 1980 (Figure 7-1, Part A). The largest decrease occurred for the sulfonylurea, tolbutamide (Figure 7-1, Part B). However, it is worth noting that the decrease began before publication of the UGDP results and that it was accompanied by increases in sales of chlorpropamide and tolazamide, also members of the family of sulfonylurea compounds.

The decline of phenformin sales, beginning in 1973, was the result of a general concern in the medical community related to isolated cases of lactic acidosis and of a negative report from the UGDP on the treatment. The drug was for all intents and purposes removed from the market in 1977 through special powers vested in the Secretary of Health, Education and Welfare (see citations 184, 185, and 187).

The de-emphasis on the oral hypoglycemic agents is reflected by advertising, as seen in the *Journal of the American Medical Association* (see Table 7-3). The only product advertised in 1979 was Pfizer's Diabinese®. In addition, advertising for the oral hypoglycemic agents represented 4.6% of the total advertising space in the journal in 1969, compared with 2.3% in 1979 (total advertising space estimated from a 25% sample of the 52 issues of the Journal published in the two time periods).

The National Therapeutic Index provides a more direct measure of physician prescribing habits. Data in this Index (IMS America, Ltd.,

5. The revised label had actually been prepared and distributed to manufacturers for use when the restraining order was issued. It contained a special warning concerning the possibility of an increased risk of cardiovascular death with the use of sulfonylurea oral hypoglycemic agents and referred specifically to the UGDP results. The label was finally revised in 1984 to include the special warning and a synopsis of the UGDP results (see citations 192, 193, and 194).

6. The National Prescription Audit is based on a nationwide sample of pharmacies that supply monthly data to IMS on the number of new and refill prescriptions issued per month. Reporting procedures were changed in 1981 and again in 1982. As a result, data obtained after the changes are difficult to compare with data obtained before the change. Therefore, they are not presented herein.

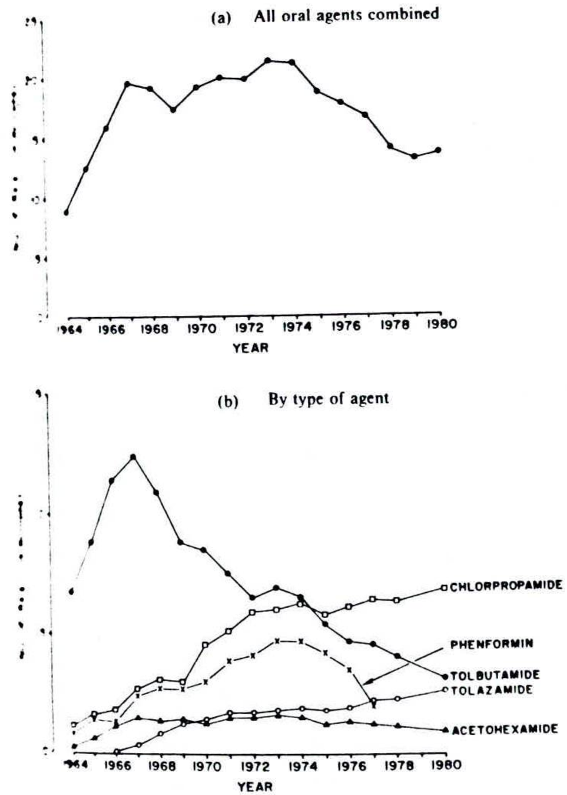


Figure 7-1 Estimated total number of hypoglycemic prescriptions (new and refill) for the U.S.

Source: Market and Prescription Data, copyright © 1964 1980, IMS America, Ltd., Ambler, Pa. (reference citation 244).

Table 7-3 Advertising for oral hypoglycemic agents in the *Journal of the American Medical Association* for 1969 and 1979

	1969		1979	
Drug	*Number of pages*	*Percent*	*Number of pages*	*Percent*
DBI®	2	1	0	0
Diabinese®	0	0	36	100
Dymelor®	11	8	0	0
Orinase®	49	36	0	0
Tolinase®	74	54	0	0
Total for hypoglycemic agents	136	100	36	100
Total number of advertising pages	2953		1597	

1977) are obtained from participating physicians. According to data in the Index, the number of physician visits of diabetics that resulted in a prescription of an oral hypoglycemic agent declined from 56% in 1969 to 36% by 1976, while the number of visits involving insulin prescriptions increased from 29% to 34% (Table 7–4). The apparent increase in use of insulin is reflected in Figure 7–2 as well. The figure suggests an increasing use of insulin relative to the oral agents. However, this conclusion is valid only if it is reasonable to assume that participating pharmacies in the National Prescription Audit have not changed their reporting habits with regard to insulin.[7]

Data from the CPHA indicate a similar trend for patient discharge data from U.S. short-term, nonfederal, general hospitals (Figure 7–3). A survey of 14 large teaching hospitals in 1969 and again in 1971 showed less reliance on oral agents and a sharper drop in their use than noted for general hospitals (Commission on Professional and Hospital Activities, 1972, 1976). The percentages of patients receiving a prescription for an oral hypoglycemic agent on discharge dropped from 33% in 1969 to 24% in 1971 for the 14 teaching hospitals, as contrasted with a drop from 38% to 34% for general hospitals. There was only a slight increase in the use of insulin for the time period in the teaching hospitals (63% in 1969 and 64% in 1971), as compared with a somewhat larger increase in the general hospitals (61% in 1969 and 64% in 1971).

The UGDP cost about $8.5 million to carry out. That cost is minuscule when contrasted with the amount of money spent on prescriptions for oral hypoglycemic agents (Table 7–5). The estimated wholesale cost of the 21 million prescriptions for oral hypoglycemic agents written in 1973 (Figure 7–1) was $105 million. This translates into an average cost of $10 per prescription, assuming retail cost is twice wholesale cost. The drop in 1980 to 13.5 million oral hypoglycemic prescriptions represents a "savings" of $75 million, about nine times the cost of the study.

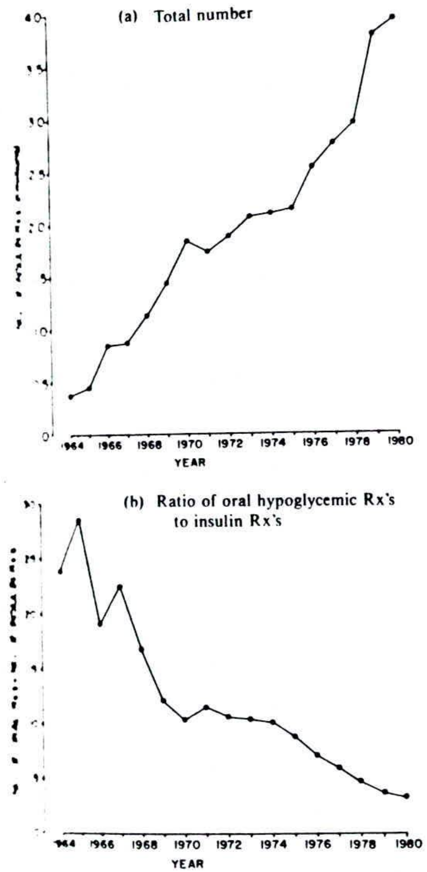
7. Technically, insulin is not a prescription drug, although it is usually issued by prescription and, hence, reported in the Audit.

Table 7-4 Percentage of patient-physician visits for diabetics by type of prescription issued

Type of Rx given	*1969*	*1970*	*1973*	*Oct. 1974 through Sept. 1975*	*Oct. 1975 through Sept. 1976*
No drug Rx	16	19	24	26	28
Drug Rx	84	81	76	74	72
Oral hypoglycemics	56	52	45	41	36
Sulfonylureas	49	45	37	34	30
Phenformin	10	11	12	10	8
Insulin	29	28	29	31	34
Total	100	100	100	100	100

Source: Market and Prescription Data, copyright © 1964–1980, IMS America, Ltd., Ambler, Pa. (reference citation 244).

Figure 7-2 Estimated number of insulin prescriptions (new and refill) and ratio of oral hypoglycemic Rx's to insulin Rx's for the U.S.



Source: Market and Prescription Data, copyright © 1964–1980, IMS America, Ltd., Ambler, Pa. (reference citation 244).

Table 7-5 Estimated U.S. wholesale dollar cost for oral hypoglycemic prescriptions

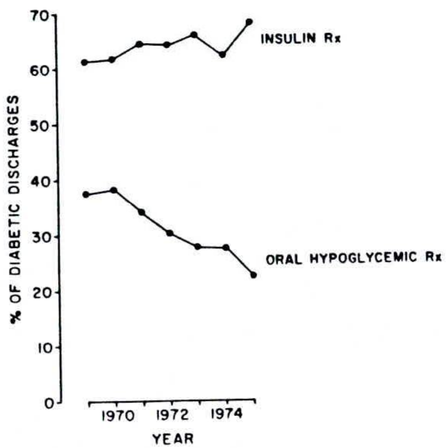
	Estimated wholesale dollar cost (in millions)*		
Year	*Phenformin*	*Tolbutamide*	*All oral agents*
1964	2.3	22.4	28.9
1965	3.9	28.2	38.6
1966	3.5	35.1	47.2
1967	7.1	38.1	58.0
1968	7.9	35.3	58.9
1969	8.4	28.7	54.5
1970	10.5	29.0	62.1
1971	14.0	24.7	65.0
1972	15.2	21.8	65.8
1973	26.7	34.8	104.8
1974	28.3	34.1	112.0
1975	26.7	31.2	109.3
1976	25.2	28.4	114.9
1977	17.1†	31.8	119.8
1978	†	30.9	109.8
1979	†	26.4	110.5

Source: Market and Prescription Data, copyright © 1964–1980, IMS America, Ltd., Ambler, Pa. (reference citation 244).

*Method of estimation changed in 1973. The large increase from 1972 to 1973 is an artifact of that change.

†NDAs for drug suspended in 1977; ordered off the market in 1978.

Figure 7-3 Type of hypoglycemic prescription on discharge from general hospitals for diabetes as a percentage of total diabetic discharges.



Source: Reference citation 82.

7.5 WAYS TO INCREASE THE IMPACT OF CLINICAL TRIALS

One obvious way to increase the impact of clinical trials is through improvements in their design and conduct. Continued proliferation of trials that have inadequate sample sizes, that involve clinically irrelevant outcome measures, and that are poorly executed cannot help having an adverse effect on the way clinical trials are viewed by the public.

The pharmaceutical industry needs to be encouraged to develop better structures for their trials. There needs to be a clearer separation of those responsible for execution of the trial from the sponsoring firm. The collection and analysis of data by firms with a proprietary interest in the product being tested is automatically open to question. Both industry and the public would ultimately benefit from trials that are above reproach.

There must also be efforts made to educate the public on the importance of clinical trials as an evaluation tool. The public must be taught to have a realistic appreciation of the strengths and weaknesses of the tool. Research societies, such as the Society for Clinical Trials and others, have a responsibility to assume leadership roles in this education process.

Investigators carrying out trials have, in effect, a public trust. They must take pains to avoid even the appearance of conflict of interest in the collection, analysis, or interpretation of results. A public trust cannot be established and maintained without high standards of integrity on the part of everyone involved in trials.

Editors of journals can help by establishing more stringent review criteria to make certain that the results of trials that are published have been generated and analyzed using sound methods. They should reject papers from trials with inadequate design features or standards of execution. Imposition of higher editorial standards would ultimately serve to elevate the design and execution standards of future trials.

Finally, as mentioned at the beginning of this chapter, there is a need for a better understanding of the way in which clinical trials influence the practice of medicine.

Part II. Design principles and practices

Chapters in This Part

The five chapters of this Part are intended to outline the primary principles and procedures to be followed in designing a trial. Chapter 8 discusses the general principles underlying selection of the study treatment, the choice of the outcome measure, and the roles of randomization and masking in data collection. Chapter 9 discusses the role of sample size and power estimates in planning a trial and details the methods for making such calculations in trials involving fixed sample size designs. Chapter 10 is devoted to a discussion of the principles and practices to be followed in administering the randomization schedule. Chapter 11 details the items that must be addressed in developing the study plan and treatment protocol for the trial. Chapter 12 outlines factors that influence the data collection schedule and contains suggestions concerning the design and content of data forms.

8. Essential design features of a controlled clinical trial

On being asked to talk on the principles of research, my first thought was to arise after the chairman's introduction, to say, "Be careful," and to sit down.

Jerome Cornfield (1959)

8.1 INTRODUCTION

The first question in any clinical trial is whether it is appropriate to mount the trial at all. Timing is of prime importance. The trial cannot proceed in the face of widespread doubts regarding its ethical base. Investigators must be satisfied that it is proper to expose patients to either the test or the control treatment. The ethical window for a trial may be quite narrow. Use of an agent in any trial setting may be deemed unethical if the agent is regarded as "too" experimental, yet that same agent may be accepted by the medical profession a short time later as the standard of treatment—without the benefit of any experimental evidence.

Ideally, the best time to start a trial is with the introduction of a treatment, before preconceived notions regarding its merit develop. Chalmers has argued for randomized trials the moment a new treatment is introduced (Chalmers, 1975, 1982b). This approach, while laudable, is not without problems. The rush to start may lead to a series of uncoordinated, small-scale efforts, none of which is adequate to answer the question of interest. Randomization of patients should not be started until there is a defined treatment protocol and a support organization to monitor the trial for evidence of adverse or beneficial treatment effects. The time involved in developing a common study protocol, writing and testing the necessary data forms, obtaining required support staff, and establishing the structure needed for proper data intake and analysis, not to mention the time needed to fund the trial, makes it difficult to start randomization with the first use of a treatment.

Once the question of timing has been resolved, the next set of issues involves basic design questions. Any controlled clinical trial requires specification of:

- A test and control treatment
- An outcome measure for evaluating the study treatments
- A bias-free method for assigning patients to the study treatments

Considerations in arriving at each of these specifications are discussed in the sections that follow.

8.2 CHOICE OF THE TEST AND CONTROL TREATMENTS

The choice of the test and control treatments is key. The general requirements to be satisfied are outlined in Table 8-1. The test treatment must be different from the control treatment; otherwise there is no point to the trial. Further, both treatments must be justifiable on medical grounds in order to allow investigators to assign patients to either treatment.

The choice of the test treatment is straightforward in settings where there is only one viable alternative to the control treatment, or where there are practical reasons for concentrating on a particular treatment (e.g., in an industry-sponsored trial done to satisfy Food and Drug Administration requirements for licensure of a particular drug). It is not when a number of

Table 8-1 Requirements for the test and control treatments

- They must be distinguishable from one another
- They must be medically justifiable
- There must be an ethical base for use of either treatment
- Use of the treatments must be compatible with the health care needs of study patients
- Either treatment must be acceptable to study patients and to physicians administering them
- There must be a reasonable doubt regarding the efficacy of the test treatment
- There should be reason to believe that the benefits will outweigh the risks of treatment
- The method of treatment administration must be compatible with the design needs of the trial (e.g., method of administration must be the same for all the treatments in a double-masked trial) and should be as similar to real-world use as practical

alternatives exist. This was the situation faced by investigators designing the University Group Diabetes Program (UGDP). They had to choose from among several different types of hypoglycemic agents (University Group Diabetes Program Research Group, 1970d). The same was true for planners of the Coronary Drug Project (CDP) in choosing among various lipid-lowering drugs (Coronary Drug Project Research Group, 1973a).

The choice of the control treatment has implications for the size of the treatment difference that can be expected. The largest difference can be expected when the control treatment is inactive. However, this design is only feasible when it is ethical to allow patients assigned to the control treatment to remain untreated (except for use of a placebo or sham treatment). The more effective the control treatment, the more difficult it will be to establish the superiority of the test treatment.

The choice of the control treatment will be dictated by current medical practice. The usual control in a surgery trial is the best available medical therapy. Some surgery trials have used sham operations as controls (Cobb et al., 1959; Dimond et al., 1960; Perry et al., 1964). However, their use has been curtailed in recent years for ethical reasons. The control treatment in a drug trial will be a standard form of drug therapy, a placebo, or no treatment at all, depending on the nature of the disease.

Treatment cannot be withheld from control patients if it is unethical to do so. Some form of medical care must be provided if a patient has a condition that requires treatment. The nature of the treatment chosen can cause a dilemma for investigators, especially when the test treatment is a refinement of the standard treatment. Investigators in the Hypertension Detection and Follow-Up Program (HDFP) had to face this problem. It was recognized that it would be unethical to identify hypertensive patients and then leave them untreated. It was also recognized that clinic personnel could not be expected to adopt two standards of care—an aggressive approach to blood pressure control for patients assigned to stepped-care and a laissez-faire approach to patients assigned to regular care. The dilemma was resolved by referring patients assigned to the control treatment back to their private physicians for treatment (Hypertension Detection and Follow-Up Program Cooperative Group, 1979b).

Some trials may involve more than one control treatment. The UGDP included both a placebo and fixed-dose insulin treatment group. The placebo treatment was used primarily for comparison with the tolbutamide and phenformin treatments, whereas both the fixed-dose insulin and placebo treatments were useful in evaluating the insulin variable treatment (University Group Diabetes Program Research Group, 1970e, 1971b, 1978, 1982).

8.3 PRINCIPLES IN THE SELECTION OF THE OUTCOME MEASURE

The outcome measure used for treatment comparisons will be a clinical event (e.g., death, myocardial infarction, significant loss of vision, recurrence of a disease) or a surrogate outcome measure (e.g., a score on a psychological test, blood pressure change, serum lipid level). The focus in this book is on trials using a clinical event as the outcome measure.

Table 8-2 provides a list of desired characteristics for the primary outcome measure. The measure should be specified when the trial is planned, before the start of data collection. Otherwise the value of the trial may be compromised, especially if there is reason to believe that data collected during the trial were used to select the measure.

The rate of occurrence of the outcome event will affect the power of the study and the length of time it is required to run (see Chapter 9). Trials involving a laboratory measure or some other surrogate outcome usually involve fewer patients and take less time to complete than those using death or some other nonfatal clinical event as the outcome, but these economies are achieved at the expense of medical relevancy. The implications of a trial with a clinical event as the outcome will, as a rule, be easier to understand than one in which clinical relevance must be inferred by relying on the presumed relationship of a surrogate outcome and the clinical condition of interest.

Table 8-2 Desired characteristics of the primary outcome measure

- Easy to diagnose or observe
- Free of measurement or ascertainment errors
- Capable of being observed independent of treatment assignment
- Clinically relevant
- Chosen before the start of data collection

It is not uncommon for trials to provide data on a number of secondary outcome measures as well. This is almost always the case in a trial in which mortality serves as the primary outcome. For example, the CDP collected data on the occurrence of myocardial infarctions and a series of other nonfatal events in addition to data on deaths (Coronary Drug Project Research Group, 1973a).

Investigators may design the trial to detect a specified treatment difference using a combination of events. Use of composite events will increase the expected event rates and hence may reduce the required size of the trial (see Chapter 9). However, the practice is ill advised because of the potential for confusion when interpreting results based on composite measures.

8.4 PRINCIPLES OF ESTABLISHING COMPARABLE STUDY GROUPS

The baseline characteristics of the test- and control-treated groups must be more or less similar in order to provide a valid basis for comparison. This need was recognized by Lind in his famous scurvy experiment. He wrote:

> *On the 20th of May 1747, I took twelve patients in the scurvy, on board the* Salisbury *at sea. Their cases were as similar as I could have them. They all in general had putrid gums, the spots and lassitude, with weakness of their knees. They lay together in one place, being a proper apartment for the sick in the fore-hold; and had one diet common to all. . . . (Lind, 1753)*

The ideal experimental model for comparing two treatments is one in which the baseline characteristics of the two study groups are identical in all aspects. This requires a homogeneous group of patients who are arbitrarily assigned to the test and control treatments. An alternative design involves enrolling pairs of patients into the trial, with each pair matched on all important baseline characteristics, and where one member of the pair is assigned to the test treatment and the other to the control treatment. However, matching is not practical. The number of patients that must be screened to find suitable matches is usually unacceptably large, to say nothing of the time required to achieve even a modest recruitment goal.

Usually the focus is on the recruitment of patients one by one, with no attempt to match. The comparability of the study groups for a few key baseline characteristics may be assured by first classifying patients into subgroups defined by those characteristics, and then assigning members of each subgroup to the test or control treatment in the same proportion as for all other subgroups. However, this approach, referred to as stratification and discussed in Chapter 10, at best can control the distribution of only a few variables.

The need for comparability can be partially satisfied by appropriate patient selection. The eligibility and exclusion criteria in most trials are designed to reduce the variability of the study populations by placing restrictions on the type of patients that may be enrolled. However, the desire for patient homogeneity and the resultant improvement in study precision must be balanced against reduced opportunities for generalizations when a highly homogeneous population is studied.

Once an eligible patient has agreed to be enrolled, it is imperative that the treatment assignment be made free of influence from both the patient and clinic personnel so as to avoid selection biases in the way study groups are formed. The general conditions that should be satisfied in order to have a sound allocation scheme are outlined in Table 8-3.

Any system in which the study physician has access to a patient's treatment assignment before enrollment is open to suspicion and violates the first requirement listed in Table 8-3. This is the main problem with allocation procedures based on characteristics associated with patients, such as birth dates or Social Security numbers. Odd-even schemes, for example, in which patients seen on odd-numbered days receive one treatment and those seen on even-numbered days receive the other treatment, are unsatisfactory for the same reason. (See Wright et al., 1954, for example.) Schemes of this sort are open to challenge and are almost always impossible to defend.

Systematic schemes in which every other patient is assigned to the test treatment violate the second requirement listed in Table 8-3. Even random allocation schemes can violate this requirement if the assignments are balanced at intervals known to clinic personnel (e.g., after every second allocation in a study involving only two study treatments). Several of the papers reviewed in Chapter 2 described or alluded to systematic nonrandom allocation schemes that appeared not to meet the second requirement (e.g., deAlmeida et al., 1980; Marks et al., 1980; Milman et al., 1980; Scott et al., 1980). However, there was not sufficient information in most of the papers to make a reliable judgment as to the soundness of the allocation process.

The third requirement, that the sequence of assignments be reproducible, is violated by any scheme that does not generate the same sequence of assignments when replicated. Coin flips are unsatisfactory for this reason, among others.

Schemes in which individual assignments are contained in sealed envelopes at the clinics are preferable to schemes described above. However, they are subject to manipulation as well if they fail to satisfy the first requirement listed in Table 8-3 (see Carleton et al., 1960, for example). Precautions must be taken to make certain that the envelopes are used in the order provided and that their contents remain unknown to clinic personnel until they are used.

The assignment process should have known mathematical properties. A major shortcoming of most informal methods of assignment, such as the odd-even scheme described above, is the absence of a mathematical base. It should also provide a clear audit trail and should be constructed and administered in such a way that departures from the established procedure can be detected.

The accepted standard for creating treatment groups is randomization. Unfortunately, there is still a good deal of misunderstanding regarding the reasons for randomizing. While the process does provide a basis for certain types of statistical analyses (Pitman, 1937), it is far more useful as a method of making bias-free treatment assignments. The term *random* is often misused in medical circles by investigators who equate haphazard and random processes (as in referring to a random blood sugar determination when really meaning a haphazard one, or in characterizing a group of arbitrarily selected individuals as a random sample). It should be reserved in research settings for processes that satisfy the definition stated in the Glossary.

Chapter 10 provides a discussion of methods for administering the treatment allocation schedule. It also contains a discussion of issues to be considered when the randomization schedule is constructed, including those related to stratification and blocking.

Table 8-3 Requirements of a sound treatment allocation scheme

- Assignment remains masked to the patient, physician, and all other clinic personnel until it is needed for initiation of treatment
- Future assignments cannot be predicted from past assignments
- The order of allocations is reproducible
- Methods for generation and administration of the schedule are documented
- The process used for generation has known mathematical properties
- The process provides a clear audit trail
- Departures from the established sequence of assignments can be detected

8.5 PRINCIPLES OF MASKING AND BIAS CONTROL

The aim of any trial should be to collect data that are free of bias, especially treatment-related bias (see Glossary for definition). The latter type of bias is of particular concern since it has the potential for obscuring a treatment difference or creating the impression that one exists when in fact it does not. The usual procedure used to protect against treatment-related bias is masking.

The term *masked*,[1] when used throughout this book, refers to a condition in which the treatment assignment, or some other item of information, is withheld from some individual or group of individuals in the study as a means of improving the objectivity of the treatment, data collection, reporting, or analysis processes. It is conventional to refer to trials as unblinded, single-blinded, or double-blinded (unmasked, single-masked, or double-masked in this book). The terms serve as decriptors of the method of treatment administration. For example, a double masked trial is one in which neither the patient nor the physician responsible for treatment is informed of the patient's treatment assignment, a single-masked trial is one in which the patient is not informed of the treatment assignment but the treating physician is, and an unmasked trial is one in which both the patient and the physician are informed of the treatment assignment. Technically, the term *single-masked* may be used to characterize a trial in which either the patient or physician is unaware of the treatment assignment; however, it usually refers to designs in which the patient is masked and the physician is not.

The logistics of masking are not simple. They are discussed in Chapter 10 in relation to bottling and dispensing drugs, and briefly in Chapter 19 in relation to data collection.

Among the randomized trials listed in the 1979 NIH Inventory of Clinical Trials (see Chapter 2), the majority were unmasked (388 out of 589, or 66%). Another 12% were single-masked, and 22% were double-masked. These results stand in marked contrast to published reports, as summarized in Chapter 2. Of the 113 trials reviewed, 76 (67%) were reported to be double-masked.

Reports of trials that are single- or double-masked should contain information on the effectiveness of the mask. The information is useful in assessing the possibility of bias in the study. However, only a few groups have addressed this issue (e.g., Beta Blocker Heart Attack Trial Research Group, 1981; Howard et al., 1981).

Treatment-specific side effects can reduce the effectiveness of the masking. This was the case with the estrogen treatments in the CDP. A total of 61% of patients assigned to the high dose of estrogen and 45% of patients assigned to the low dose of estrogen complained of decreased libido, as contrasted with only 1.5% of the placebo-treated patients (Coronary Drug Project Research Group, 1970a).

The principle of masking is general and applies whenever it is practical to withhold information that, if known, may influence the way in which data are collected or how the treatments are adminstered. Table 8-4 lists suggested masking guidelines. Masked data collection is especially important in trials involving outcome measures that are subject to measurement or ascertainment errors.

A double-masked trial, as defined above, is characterized by masked data collection. However, even single-masked or unmasked trials may be designed so that data collection is done in a masked fashion via structures in which treatment information is withheld from personnel responsible for data collection. The structures require one set of personnel to administer the study treatments and another to collect the data needed for assessing the study treatments.

Laboratory tests should be performed, recorded, and reported by personnel who are masked to treatment assignment, regardless of the level of treatment masking in the rest of the trial. The only exceptions are cases in which treatment assignment is needed to determine the tests to be performed. Likewise, records such as ECGs, fundus photographs, and x-rays should be read by individuals who are masked to treatment assignment. The same is true of personnel responsible for coding or classifying outcome events.

Ideally, all keying, editing, and data analysis activities in the data center should be performed by personnel who are masked. This standard is not easy to achieve because of the obvious prac-

Table 8-4 Masking guidelines

- Use a treatment allocation scheme that meets the masking criteria listed in Table 8-3 (i.e., the treatment assignment for a patient cannot be determined in advance of enrollment)
- Administer treatments with the highest level of masking feasible (e.g., double-masked if possible; single-masked if double masking is impossible; unmasked only if any level of masking is out of the question)
- Require, when possible, that essential data collection, measurement, reading, and classification procedures on individual patients be made by persons who have no knowledge of treatment assignment or course of treatment
- Require, when possible, that outcome measurements that are subject to interpretation errors (e.g., measurements requiring a subjective evaluation) be made by personnel who are masked to treatment assignment
- Do not require masked treatment administration if doing so requires study patients to assume measurable risks in order to achieve or maintain the masking

1. Used in this book instead of *blinded* because it is regarded as a more apt description of the process involved. Further, use of the latter term, as in double-blinded trial, leads to confusion in some settings, such as in vision trials where the outcome measure is blindness.

tical problems involved in maintaining the mask. However, it is important when it cannot be achieved to make certain that decisions regarding the way in which data are keyed or used for analysis purposes are made without regard to treatment assignment or observed treatment differences.

The principle of masking has been extended to treatment monitoring committees as well. Treatment monitoring reports presented in the Diabetic Retinopathy Study (DRS) were masked with regard to treatment group, even though the trial itself was unmasked. However, in this case the masking was subsequently abandoned because of the logistical difficulties involved in producing the monitoring reports and because of its limited usefulness (Knatterud, 1977).

9. Sample size and power estimates

A difference to be a difference must make a difference.

Source unknown

9.1 SEQUENTIAL VERSUS FIXED SAMPLE SIZE DESIGNS

This chapter deals with sample size and power estimates for fixed sample size designs. All of the trials sketched in Appendix B are of this type. Strictly speaking, a fixed sample size design is one in which the investigator specifies the required sample size before starting the trial. The specification may be based on a formal sample size calculation or on practical considerations related to cost, patient availability, or other factors. The investigator then proceeds to enroll the number of patients specified, unless there are extenuating circumstances to the contrary (e.g., the specified number cannot be recruited as planned or recruitment has to be stopped because of adverse or beneficial treatment effects). In practice, the sample size may not be set until after the trial is started or may never be formally set in some cases. In other cases, it may be merely implied by other conditions, such as the amount of time allowed for patient recruitment.

The approach is quite different with sequential designs. A classical open sequential design provides for continued patient enrollment until the observed test-control treatment difference exceeds a predefined boundary value (see Figure 9-1). The simplest application of this design is the enrollment of patients in pairs. One member of each pair is assigned to the test treatment and the other member is assigned to the control treatment. The decision as to whether to enroll the next pair of patients is based on outcomes observed for patients already enrolled. The pair of patients is enrolled if the cumulative test-control difference for all previously enrolled pairs of patients is still within the defined boundaries. The pair is not enrolled if one of the two boundaries is exceeded.

The expected sample size, given a specified type I and II error level, is smaller for a sequential design than for its fixed sample size counterpart (Armitage, 1975). However, the number of patients required in any given replication can exceed the number required with a fixed sample size design. In fact, there is a chance, albeit infinitesimally small, that the treatment difference will remain within the defined boundaries no matter how many pairs of patients are enrolled. This possibility is eliminated by imposing a limit on the number of patients that may be enrolled, as illustrated in Figure 9-2. Closed sequential designs (so named because of the limit imposed on the number of patients that may be enrolled) are preferred to open sequential designs in most medical settings because they allow the investigator to stop the trial if the study treatments appear to be of about equal value.

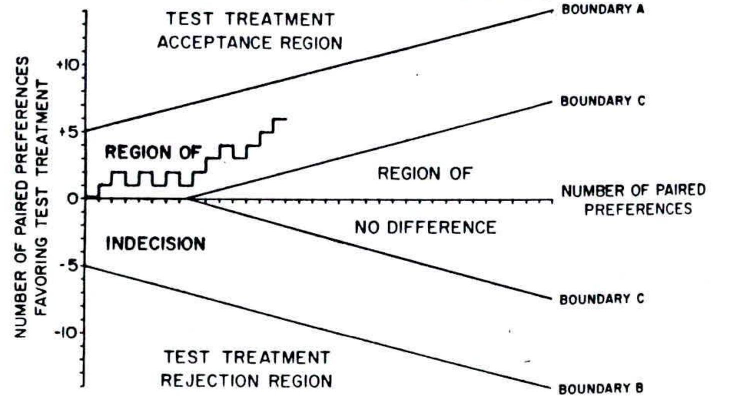
The initial work on open sequential designs was done by Wald (1947). The closed modifications come from work by Bross (1952) and Armitage (1957). A book by Armitage (1975) focuses on applications of closed designs to medical trials. (See Grant, 1962, and Snell and Armitage, 1957, for examples of the two types of sequential designs).

There are sequential aspects to any trial, even one done using a fixed sample size design. Patients in both types of designs are typically enrolled over time. The temporal nature of the enrollment process leads to a gradual accumulation of outcome data for use in making treatment comparisons. As noted above, a new comparison is made after each pair of patients is enrolled in the classical sequential design. The results of the comparison are used to decide whether to stop patient enrollment. The decision-making process is more complicated in the typical fixed sample design, at least for the class of trials discussed in this book. An investigator must not only decide whether it is appropriate to continue patient enrollment up to the limit set, but also whether it is appropriate to continue the trial after enrollment is completed. He should stop the trial once it becomes clear that the test treatment is superior or inferior to the control treatment, regardless of whether patients are still being enrolled (see Chapter 20 for further discussion).

The use of sequential designs is limited to situations in which outcome assessment can be made shortly after patients are enrolled in the trial. They are not practical where long periods of follow-up are required to accumulate sufficient outcome data to make reliable treatment comparisons. The usual approach in such cases is to use a fixed sample size design. This approach, as discussed herein, utilizes a frequentist analysis philosophy—a philosophy based on work of Neyman and Pearson (1966) and one that is widely used in biostatistics for analysis of medical research. Other analysis philosophies include those built on the likelihood principle and on Bayes' theorem. Plackett (1966) has reviewed all three philosophies. The frequentist approach is reviewed by Armitage (1963) and by Armitage and co-workers (1969). The likelihood approach is reviewed by Anscombe (1963). The Bayesian approach is reviewed by Colton (1963) and Cornfield (1966a).

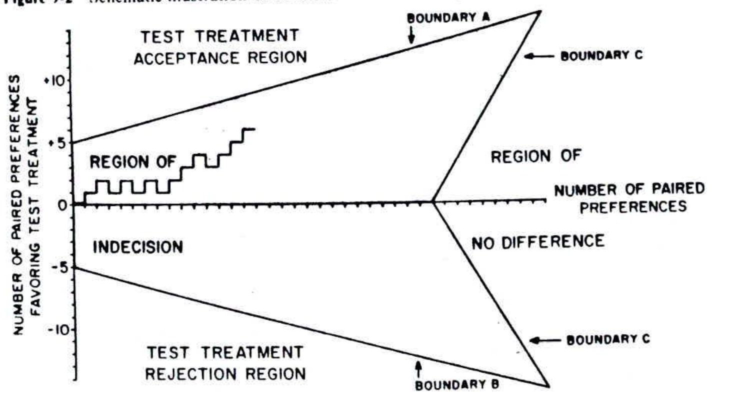
Sample size and power estimates for fixed sample size designs are discussed by a number of

Figure 9-1 Schematic illustration of boundaries for open sequential design.



Note: Trial continues until observed number of preferences (ignoring ties) crosses a boundary line. The test treatment is considered superior to the control treatment if boundary line A is crossed, inferior to the control treatment if boundary B is crossed, and equal to the control treatment if boundary C is crossed. The C boundary lines are deleted in trials designed to continue until the test treatment is declared superior or inferior to the control treatment

Figure 9-2 Schematic illustration of boundaries for closed sequential design.



Note: Trial continues until observed number of preferences (ignoring ties) crosses a boundary line. The test treatment is considered superior to the control treatment if boundary line A is crossed, inferior to the control treatment if boundary line B is crossed, and equal to the control treatment if boundary line C is crossed.

authors, including Cochran and Cox (1957), Cox and Hinkley (1974), Fleiss (1981), Lachin (1981), Schlesselman (1982), and Snedecor and Cochran (1967). Readers may refer to these references or to other basic statistics texts for details not covered in this chapter.

9.2 SAMPLE SIZE AND POWER CALCULATIONS AS PLANNING GUIDES

It is unwise to undertake a fixed sample size trial without a calculation to determine the number of patients required or the power available with a specified sample size. With a sample size calculation, the investigator sets out to determine the number of patients required to detect a designated treatment difference with specified levels of type I and II error protection. With a power (see Glossary) calculation, the investigator determines the power associated with a specified treatment difference, given a specified sample size. Either one of these calculations, may lead to subsequent design modifications. The modifications may include expansion from a single center to multiple centers to increase the number of patients available for study, changes in the patient admission criteria to make recruitment easier, or abandonment of the trial.

The archives of clinical trials are cluttered with inconsequential trials. Such trials are, in one sense, unethical in that they require patients to accept the risks of treatment, however small, without any chance of benefit to them or future patients. Small-scale preliminary investigations may be justified when part of a larger plan, but not as an end in their own right.

The absence of a planned approach to study design is evident from a review of the published literature, as discussed in Chapter 2. Few of the trials cited there show any evidence of having involved sample size or power calculations (see also Freiman et al., 1978, and Mosteller et al., 1980).

The design documents prepared when the trial is planned should indicate the recruitment goal for the trial and how it was determined. If the goal was the result of a sample size calculation, the details of that calculation should be provided. If it was set by practical considerations, such as cost or the presumed availability of patients, it should be accompanied by appropriate calculations to indicate the power that can be expected with the proposed number of patients.

In either case, the calculations, such as shown in Tables 9-1 and 9-2, should indicate how the trial is affected if the control event rate used in the sample size calculation proves to be wrong or how power changes as a function of sample size.

The main thrust of the discussion in this chapter relates to the use of sample size and power estimates in planning the trial. However, as noted in Section 9.7, the same methods are used for sample size adjustments during the trial or for posterior power calculations at the end of the trial.

9.3 SPECIFICATIONS FOR SAMPLE SIZE CALCULATIONS

A determination of the required sample size cannot be undertaken until the basic design features of the trial, such as outlined in Table 9-3, have been set. It may take months and a good deal of interaction between investigators, especially between the physicians and biostatisticians, to reach agreement on the specifications.

The subsections that follow detail the considerations that go into setting the specifications and provide discussion of the ways in which they influence sample size requirements. Most of the same points pertain to power calculations as well.

9.3.1 Number of treatment groups

The considerations involved in reaching a decision on the type and number of study treatments have been discussed in Chapter 8. The sample size formulations presented in Section 9.4 are for the case of a trial involving one test and one control treatment. However, they can be used for trials with any number of test treatments, so

Table 9-1 Illustration of a sample size presentation, $\alpha = 0.01$ (two-tailed), $\beta = 0.05$, and $\lambda = 1$

Control event rate, P_c	$\Delta_R = \frac{P_c - P_t}{P_c}$ *0.125*	*0.250*	[illegible]
0.10	19,373	4,549	[illegible]
0.15	12,246	2,886	[illegible]
0.20	8,682	2,055	[illegible]
0.30	5,119	1,223	[illegible]

Table 9-2 Illustration of a power presentation, given a sample size of 800, $\alpha = 0.01$ (two-tailed), and $\lambda = 1$

Control event rate, P_c	$\Delta_R = \frac{P_c - P_t}{P_c}$ *0.125*	*0.250*	*0.375*
0.10	0.043	0.209	0.432
0.15	0.668	0.359	0.192
0.20	0.968	0.476	0.067
0.30	0.181	0.201	0.004

long as investigators plan to allocate the same number of patients to each of the test treatments. The total sample size, N, for a trial with one control treatment and uniform allocation among the r test treatments is:

$$N = rn_t + n_c \tag{9.1}$$

where

r = number of test treatments

n_t = sample size required for each of the r test treatments

and

n_c = sample size for the control treatment

For the purposes of calculation it is necessary to specify the test-control allocation ratio,

$$\lambda = n_t/n_c \tag{9.2}$$

Table 9-3 Design specifications affecting sample size considerations

- Number of treatment groups to be studied
- Outcome measure
- Anticipated length of patient follow-up
- Alternative treatment hypothesis
- Detectable treatment difference
- Desired type I and II error protection
- Allocation ratio
- Anticipated rate of loss to follow-up
- Anticipated treatment noncompliance rate
- Anticipated treatment lag time
- Degree of stratification for baseline risk factors
- Level of type I and II error adjustment for multiple comparisons
- Level of type I and II error adjustment for multiple looks
- Level of type I and II error adjustment for multiple outcomes

It is simply the number of patients to be assigned to a test treatment divided by the number to be assigned to the control treatment. This quantity is fixed by the study investigators and is generally the same for each of the r test treatments in the trial (see Section 9.3.7 for factors determining the choice of λ).

The usual approach is to calculate the sample size requirement for n_t using the formulas given in Section 9.4 and then to derive the value of n_c from Equation 9.2 by noting that $n_c = n_t/\lambda$. For example, given $r = 3$, $\lambda = 0.5$, and a calculated value of 50 for n_t yields an n_c of $50/0.5 = 100$. The total sample size is 250, as derived from Equation 9.1.

The sample size is not given by Equation 9.1 if the trial involves more than one control treatment. The simplest approach in this case is to use just one of the control treatments for the sample size calculation—ideally the one that provides the best basis for assessing the test treatments. The value for n_c, as derived from Equation 9.2, would be used for each control treatment and the total sample size would be $N = rn_t + sn_c$, where s is the number of control treatment groups. An alternative approach, if a minimal level of type II error protection is desired for comparisons of a test treatment with any one of the control treatments, involves making a sample size calculation for each of the different control treatments and then using the largest value of N obtained to plan the trial.

9.3.2 Outcome measure[1]

Sample size and power formulations are given in this chapter for binary as well as continuous outcome measures. However, the main emphasis is on binary outcomes (see Glossary) because of the class of trials considered in this book, as outlined in Chapter 1. Trials with binary outcomes are characterized by data collection schemes in which patients may be classified at any point after enrollment as either having or not yet having experienced the event of interest. The event may be a desired or undesired outcome depending on the trial and patients selected for study. It will be desired (positive) in trials in which patients are watched for disappearance or amelioration of some medical condition. It will be undesired (negative) when they are watched for the occurrence of death or some

1. See Chapter 8 for additional discussion of factors influencing the choice of the primary outcome measure.

morbid event. All discussion and calculations in this chapter are for negative events.

For the purposes of sample size or power calculations, the investigator may decide to alter the form of the outcome measure when the underlying measure is polychotomous or continuous. The choice should be dictated by the anticipated analysis requirements at the end of the trial. The calculations should be made using the unaltered underlying measure if the aim of the trial is to assess distributional changes in the measure over the time course of the trial. They should be made using a binary event measure, constructed from a dichotomization of the underlying measure, if observed values of the measure, at or above a specified level, take on special medical or operational significance (e.g., as in the case with blood pressures over a defined level used to diagnose hypertension and signal the need to initiate treatment).

The decision as to whether to design the trial to detect a mean change in some continuous measure or a difference in some event rate can have major implications on how the trial is perceived when it is finished. It is one thing to conclude that there is a significant difference in mean diastolic blood pressure between the study treatment groups, and quite another to conclude there is a significant difference in the rate of development of hypertension in the two groups. The latter statement has far greater clinical relevance than the former.

9.3.3 Follow-up period

Sample size and power calculations require specification of the follow-up period. Generally, the longer the period, the higher the accumulated event rate and the smaller the required sample size for a given type I and II error level.

The specification used for planning purposes may be modified as the trial proceeds. For example, the follow-up period may be extended to compensate for a shortfall in patient recruitment or for a lower-than-anticipated control event rate. Or it may have to be curtailed because of funding or other problems.

9.3.4 Alternative treatment hypothesis

The calculations in this chapter are always made under the null hypothesis of no treatment effect versus a specified alternative. The alternative will be constructed to cover treatment effects of a specified size that are either beneficial or adverse (two-sided alternative), or that are only beneficial (one-sided alternative). The decision as to whether to use a one- or two-sided alternative depends on the clinical importance of a positive versus a negative treatment effect and how much is known about the safety of the test treatment when the trial is planned.

Trials of the sort considered in this book are done to establish the efficacy of a treatment, not toxicity. This fact argues for use of a one-sided alternative in the calculation, even though practice seems to favor use of two-sided alternatives. The reason has to do with the amount of prior evidence investigators have regarding treatment safety. They may prefer a two-sided alternative simply as a means of documenting their own uncertainty regarding the potential merits of the test treatment. A side benefit of the practice is that it leads to a larger sample size than the same α and β using a one-sided test. The increased sample size represents a hedge against unanticipated losses, such as those due to lack of compliance to the treatment protocol during the trial.

9.3.5 Detectable treatment difference

The experimenter is required to specify the minimum treatment difference he wishes to detect under the alternative hypothesis. The larger the difference the smaller the required sample size.

The difference chosen should be realistic. A 50% reduction in the test treatment event rate, while of unquestioned clinical relevance if achieved, is unlikely in real life. Only miracle treatments produce reductions of this size and there are few such treatments around and even fewer that require discovery via a clinical trial. Generally, the gains with most new treatments are much more modest. Certainly a reduction in the event rate does not have to be enormous to be important. Small reductions, in the range of 5 to 10%, can have major public health implications if they apply to death or some other serious nonfatal event associated with a common disease.

9.3.5.1 *Binary outcome measures*

Specification of the difference, in this case, requires the experimenter to designate a value for both P_c and P_t, or for P_c and the percentage reduction in P_c to be achieved with the test treatment,

where

P_c = anticipated event rate for the control-treated group

and

P_t = anticipated event rate for the test-treated group

The minimal detectable difference expressed in absolute terms is:

$$\Delta_A = P_c - P_t \tag{9.3}$$

Expressed in relative terms, it is:

$$\Delta_R = \frac{P_c - P_t}{P_c} \tag{9.4}$$

Although either form is acceptable, many investigators prefer to express the difference in relative terms using Δ_R, even though sample size and power formulas are conventionally expressed in terms of Δ_A. Equation 9.5 can be used to convert from a relative to an absolute difference

$$\Delta_A = P_c \Delta_R \tag{9.5}$$

Ideally, the value chosen for P_c should be derived from follow-up studies of patients similar to those to be enrolled in the trial and who received treatment similar to that planned for the control-treated group. Unfortunately, follow-up data such as these are usually not available. Hence, the experimenter may have to rely on an educated guess for P_c. The value chosen may turn out to be higher or lower than the one actually observed in the trial. Selection of a value for P_c that is lower than the one subsequently observed means the trial was larger than it needed to be to detect a given relative difference not a serious problem unless the underestimation resulted in a significant increase in the cost and time needed to carry out the trial. (The reverse is true if $P_c + P_t < 1$ and the experimenter is interested in detecting a prespecified absolute difference.) A more serious and common problem stems from overestimation of P_c. The sample size estimate in such cases will be smaller than needed to achieve the required error protection. Overestimation can occur even in instances in which investigators have reasonably reliable information for determining P_c, as in the Coronary Drug Project (CDP). The P_c (5-year mortality rate) used for sample size calculations was 30 per 100 population. The observed rate was only 21 per 100 population (Coronary Drug Project Research Group, 1975).

The tendency to overestimate P_c arises, at least in part, from the failure or inability of the study planners to predict the impact of the proposed patient eligibility criteria on subsequent observed event rates. The exclusion of seriously ill patients from enrollment may well yield a population with a better than expected prognosis.

9.3.5.2 *Continuous outcome measures*

The difference to be detected in this case is expressed as a function of means, as discussed in Section 9.4.2. The variance estimate[2] required, like the value of P_c used for binary outcomes, should be based on actual data if at all possible. It is wise to explore the effect of a range of variance estimates on sample size if there is no reliable way of estimating variance before the start of the trial.

9.3.6 Error protection

The choice of α and β (probabilities of type I and II error, respectively, see Glossary for definition) is arbitrary. The first instinct of an inexperienced investigator is to want a trial that precludes any possibility of either type of error—a lofty goal since an infinite number of patients is required to achieve it!

The choice of α and β should depend on the medical and practical consequences of the two kinds of errors. Relatively high error rates (e.g., $\alpha = 0.10$ and $\beta = 0.2$) may be acceptable for preliminary trials that are likely to be replicated, whereas lower rates (e.g., $\alpha = 0.01$ and $\beta = 0.05$) should be used if replication is unlikely.

The consequences of both a type I and II error must be considered. For example, one might choose:

- $\alpha = \beta$ if both the test and control treatments are new, about equal in cost, and there are good reasons to consider them both relatively safe
- $\alpha > \beta$ if there is no established control treatment and the test treatment is relatively inexpensive, easy to apply and is not known to have any serious side effects
- $\alpha < \beta$ if the control treatment is already widely used and is known to be reasonably safe and effective, whereas the test treatment is new, costly, and produces serious side effects

2. The need for an independent variance estimate is avoided for binary outcomes. The variance in such cases is a function of the specified event rates.

The most common approach is to set $\alpha < \beta$. However, this is only reasonable if the consequences of a type II error are considered to be less than those of a type I error.

9.3.7 Choice of allocation ratio

The allocation ratio is ordinarily under the control of the experimenter and is set before patient enrollment is started, except with some forms of adaptive allocation (see Chapter 10). All of the trials sketched in Appendix B involved preset allocation ratios, except one (see item 20.a of Table B-4 in Appendix B). A uniform allocation scheme, in which the probability of assignment to one treatment group is the same as for all other treatment groups, is generally preferred (used in 11 of the 14 trials sketched in Appendix B; see item 20.e of Table B-4). Nonuniform methods of allocation are used when there is a need to concentrate more patients in certain treatment groups to satisfy secondary aims of the trial or to provide increased precision for certain of the treatment comparisons. Investigators in the Persantine Aspirin Reinfarction Study (PARIS) decided to allocate twice as many patients to the aspirin and to the aspirin-persantine treatment groups than to the placebo treatment group. They made this choice because they considered comparison of the aspirin and aspirin-persantine treatment groups more important than comparison of either of these treatment groups with the placebo treatment group (Persantine Aspirin Reinfarction Study Research Group, 1980b).

The CDP allocated 2.5 times as many patients to the placebo treatment as to any one of the five test treatments (Coronary Drug Project Research Group, 1973a).[3] The allocation ratio was chosen to minimize the variance for the five test-control comparisons of 5-year mortality. A somewhat lower ratio would have been derived using an approach developed by Dunnett (1955). His method assumes the experimenter wishes to construct confidence intervals about each test-control outcome difference at the end of the trial, such that the risk of a type I error for all comparisons combined is α. This error probability condition is satisfied if $\sqrt{r}$ patients are assigned to the control treatment for every patient assigned to any one of the r test treatments used in the trial. The CDP would have required an allocation ratio of 2.24 instead of the one used with this method of calculation.

9.3.8 Losses to follow-up

Losses to follow-up are concentrated in drop-outs (throughout this book, patients enrolled in the trial who are no longer able or willing to return to the clinic for regular follow-up examinations) who can no longer be followed for the event of interest. A patient who drops out is automatically lost to outcome follow-up unless the event used for outcome assessment can be reliably observed and reported outside the clinic setting.

The loss to follow-up rate used in the sample size calculation is estimated by the experimenter. As with other variables, such as P_c, the value used should be based on relevant experience if at all possible. The value chosen may be zero or very small in cases in which it is possible to continue follow-up of patients for the outcome of interest even if they refuse to return to the clinic for regular follow-up examinations (e.g., as with patients under follow-up for mortality or some other event that can be reliably observed and recorded outside the clinic setting). It may have to be set quite high if long periods of clinic surveillance are required for outcome measurement.

Clearly, any loss of outcome data, regardless of how it occurs, will reduce the statistical precision of the trial, and may introduce bias as well if the losses are differential by treatment group. Hence, the sample size estimate, as given by formulas in Section 9.4, must be increased to compensate for the anticipated loss. This is normally done by multiplying the sample size by the quantity, $1/(1 - d)$, where d is the anticipated loss rate. For example, a d of 10/100 would mean that for every 100 patients enrolled, 10 could not be classified as to presence or absence of the outcome of interest because of the lack of follow-up data. It would require multiplying the calculated sample size by a factor of $1/0.9 = 1.11$ to compensate for the losses.

9.3.9 Losses due to treatment noncompliance

The sample size must be increased to compensate for loss of precision due to treatment noncompliance as well. Treatment compliance is rarely an all-or-none phenomenon. The level of compliance achieved may range from low to high, depending on the patient. Perfect compliance may be difficult, if not impossible, to achieve, especially in drug trials where the patient is required to take the assigned medication over long periods of time.

There are two aspects to the determination of compliance. One has to do with the amount of exposure the patient has to the assigned treatment, and the other has to do with the amount of exposure to the other study treatments. Underexposure to the assigned treatment may arise from

- Patient unwillingness to accept the assigned treatment
- Physician unwillingness to administer it
- Patient or physician unwillingness to use the full treatment dosage

Overexposure to the assigned treatment may arise from:

- A mistake by the study physician or patient in the assigned treatment (e.g., as in the case in which a patient takes twice as many pills as required)
- Administration of the same treatment outside the study clinic by the patient's private physician
- Patient self-treatment with medications obtained outside the study clinic (e.g., as with a patient in a myocardial infarction trial who is assigned to aspirin therapy and who takes his medication but who uses his own supply of aspirin for headaches and other ailments as needed)

Exposure to one of the other study treatments may arise in various ways. Examples include:

- A patient who takes a drug outside the trial that is similar to one of the test treatments (e.g., as with a patient assigned to the control treatment in an aspirin trial who uses an over-the-counter cold remedy containing aspirin)
- A patient who demands and receives, midway in the course of the trial, another study treatment in place of, or in addition to, the assigned treatment
- A physician who unwittingly switches a patient from the assigned treatment to another study treatment through a mix-up in prescriptions
- A physician who elects, for medical reasons, to administer another study treatment to a patient in the trial, in addition to, or in place of, the patient's assigned treatment

Any departure from the study treatment protocol, regardless of the nature of the departure, reduces the chances of finding a treatment difference. For example, a patient assigned to the test treatment who refuses the treatment may, in effect, expose himself to the control treatment (e.g., as in the case where the control treatment involves no treatment at all). This reduces the chance of finding a treatment difference, even if the adherence of patients actually assigned to the control treatment is excellent. Conversely, so does the exposure of control-treated patients to the test treatment (e.g., as in a coronary bypass surgery trial where a sizable number of the control-treated patients receive bypass surgery), even if the compliance of patients assigned to the test treatment is excellent.

Loss of precision due to noncompliance is not necessarily related to patient follow-up status. Dropouts, to be sure, automatically become noncompliant if the treatments to which they were assigned are stopped when they drop out. However, as noted above, patients who do not drop out can become noncompliant as well. Further, being a dropout does not necessarily imply a state of noncompliance if the treatment process, as specified by the study protocol, was completed before dropout and the patient is not exposed to any of the study treatments after dropout.

The loss of precision due to noncompliance is compensated for in the same way as losses to follow-up, as discussed in Section 9.3.8. The value for d will be based on the amount of noncompliance anticipated and its role in reducing the precision of the trial. Trials with losses from follow-up and noncompliance will require a composite multiplier to account for both kinds of losses. For example, the CDP used a combined d of 0.30. In actual fact, the losses were due almost exclusively to noncompliance, since it was possible to follow virtually every patient for mortality—the primary outcome measure.

9.3.10 Treatment lag time

Most calculations are made as if there is no treatment lag (i.e., the full effect of the treatment is realized as soon as it is applied). That convention is followed in this chapter. The approach is

3. The actual ratio, as derived by methods described in the 1973 CDP publication, was 2.45. It was rounded up to 2.5 to simplify construction of the allocation schedules needed in the trial.

reasonable with some forms of treatment (e.g., most types of surgery and certain drug treatments), but not for others (e.g., with a drug given to dissolve atherosclerotic plaques). The decision of investigators in the Anturane Reinfarction Trial to ignore deaths that occurred within seven days after the initiation of treatment was based on a presumed treatment lag for Anturane (Anturane Reinfarction Trial Research Group, 1978, 1980; Temple and Pledger, 1980). One reason for ignoring lag times has to do with the mathematical difficulties involved in taking account of them in sample size calculations. Further, there is often no reliable way to estimate lag times.

The impact of lag time on sample size is illustrated below for a trial involving 5 years of follow-up for each patient enrolled, one test and one control treatment, $\lambda = 1$, $P_c = 0.30$, $\Delta_R = 0.30$, $\alpha = 0.01$ (one-tailed), $\beta = 0.05$, and $d = 0$. The sample sizes recorded are derived from tables developed by Halperin and coworkers (1968). The required sample size, given the above specifications, is 1,474 if the full effect of the treatment is realized as soon as it is administered. It is about twice this size if it takes 2.5 years for the treatment to reach full effectiveness and nearly 20 times as large if the lag time is 10 years.

Lag time	*Sample size:* $n_t + n_c$	*Sample size ratio**
0	1,474	1.00
2 months	1,530	1.04
6 months	1,656	1.12
1 year	1,870	1.27
2 years	2,444	1.66
2.5 years	2,828	1.92
3 years	3,266	2.22
4 years	4,492	3.05
5 years	6,536	4.43
7.5 years	12,136	8.23
10 years	29,428	19.96

*Ratio of sample size for indicated lag time relative to size for 0 lag time.

9.3.11 Stratification for control of baseline risk factors

The sample size is influenced by the amount of stratification done to control for baseline variation. The issues involved in selection of stratification variables are discussed in Chapter 10. Technically, the sample size calculation should take account of the stratification planned. However, in actuality, most calculations are made ignoring stratification. Doing so can lead to an overestimate of the required sample size if the variables used for stratification represent important risk factors and if the calculated sample size is small (see Section 10.3.2 of Chapter 10).

9.3.12 Degree of type I and II error protection for multiple comparisons

The experimenter must also decide whether the error protection specified is to be for a single treatment comparison or for multiple treatment comparisons (see Glossary and Section 20.4 of Chapter 20). Section 9.3.7 alludes to methods of sample size calculation in which the investigator is interested in r test-control treatment comparisons. However, the need for making multiple comparisons is not limited to such cases. It can be just as great when $r = 1$ (i.e., where the trial involves only two treatment groups) if the investigator wishes to design the trial to provide a specified level of error protection for treatment comparisons within designated subgroups of patients. One approach in this setting is to calculate sample size estimates for each subgroup of interest. A drawback with it is that it leads to a series of recruitment quotas—one for each subgroup (see Section 14.1 of Chapter 14). An alternative and generally preferable approach is to ignore subgroups in making the sample size calculation and then to estimate the power provided for subgroups of interest. The total sample size may be increased (e.g., by making a new calculation using smaller values for α and β) if the power is considered to be inadequate for one or more of the subgroups of interest.

9.3.13 Degree of type I and II error protection for multiple looks for safety monitoring

The experimenter may plan to look at outcome data at various time points over the course of the trial in conjunction with the safety monitoring process (see Glossary and Chapter 20). Carrying out multiple looks will alter the type I and type II error levels (see Dupont, 1983a). Ideally, sample size should be estimated with the need for safety monitoring in mind from the

onset. However, calculations are routinely made ignoring the need, in part because of difficulties in making the necessary adjustments. This practice is followed here. However, designers of trials should recognize that the error protection provided in such cases will be less than the levels used in making the calculations.

9.3.14 Degree of type I and II error protection for multiple outcomes

A trial, even though planned to focus on a primary outcome, will generate data for a number of secondary outcomes as well (see Glossary for definitions of primary and secondary outcome measures). The usual approach is to base the sample size calculations on the primary outcome of interest and to accept whatever power that calculation yields for the comparisons involving secondary outcome measures.

The only compensation made may be in the choice of α and β. The investigator may choose smaller values than normally used as a means of increasing the precision for the primary as well as secondary comparisons. He may be forced to make calculations for each outcome and then to use the largest size for planning the trial if he is unwilling to designate any of the measures as primary.

9.4 SAMPLE SIZE FORMULAS

Table 9-4 provides a summary of the calculations discussed in this Section and Section 9.5. Tables 9-5 and 9-6 are included for use in making sample size and power calculations. Other more extensive tables of the two functions may be found in many texts on statistics.

The method of analysis implied in the sample size calculation should be identical to that used when the results of the trial are analyzed. However, this is not always possible, as already noted with regard to the need for safety monitoring and the use of secondary outcomes in the analysis process. Technically, there are as many methods of sample size calculation as there are methods of data analysis. The methods presented in this Section are the most common ones.

The methods presented assume that the primary comparison will entail a simple comparison of proportions constructed at the end of the trial (or after a specified period of patient follow-up). Strictly speaking, they are not appropriate if the treatment groups are to be compared using life-table methods. The log rank test is the test of choice in such cases (Gail, 1985). It will yield smaller sample sizes than are obtained with the tests covered herein (i.e., it is more efficient). The difference is small for trials involving rapid patient accrual and low event

Table 9-4 Sample size and power calculation summary for Sections 9.4 and 9.5

Test	*Sample size*	*Power*	*Assumptions*	*Applicability*
A Binary outcome measure				
Fisher's exact test	See Section 9.4.1.1	See Section 9.5.1.1	Independent observations	Applicable over entire event rate range from 0 to 1
Chi-square approximation	Eqs 9.6, 9.7	Eqs 9.16, 9.17	Independent observations	P_c and $P_t \geq 0.2$ but ≤ 0.8; n_cP_c, n_cQ_c, n_tP_t, and n_tQ_t all ≥ 15
Inverse sine transform approximation	Eqs 9.8, 9.9	Eqs 9.18, 9.19	Independent observations	P_c and $P_t \geq 0.05$ but ≤ 0.95; n_cP_c, n_cQ_c, n_tP_t, and n_tQ_t all ≥ 15
Poisson approximation	Eqs 9.10, 9.11	Eqs 9.20, 9.21	Independent observations	Low event rates (e.g., P_c and $P_t \leq 0.05$; n_cP_c, and $n_tP_t \geq 10$)
B Continuous outcome measure				
Normal approximation for 2 independent means	Eqs 9.12, 9.13	Eqs 9.22, 9.23	Independent observations Common variance Normality	n_c and $n_t \geq 30$
Normal approximation for mean change	Eqs 9.14, 9.15	Eqs 9.24, 9.25	Independent observations between patients Common variance Normality	n_c and $n_t \geq 30$

Table 9-5 Z values for $N(0, 1)$ distribution for selected error levels

Error level	*One-tailed*	*Two-tailed*
0.500	0.000	0.674
0.400	0.253	0.842
0.300	0.524	1.036
0.200	0.842	1.282
0.100	1.282	1.645
0.050	1.645	1.960
0.025	1.960	2.248
0.010	2.326	2.576
0.005	2.576	2.813

rates. It is largest for trials involving slow accrual and high event rates.

All of the formulations in this chapter are for one-tailed tests. However, they may be used for two-tailed tests by using $\alpha/2$ wherever α appears in the formulas cited. Strictly speaking this substitution should be used only when the allocation ratio, λ, equals 1, since there are disagreements among statisticians as to the validity of the substitution when $\lambda \neq 1$. However, the common practice is to use the substitution even if $\lambda \neq 1$.

9.4.1 Binary outcome measures

9.4.1.1 Fisher's exact test

Fisher's exact test is the test of choice for comparing simple counts or proportions based on binary data (Gart, 1971). The test, unlike others considered in this section, works for samples of any size. It yields an exact p-value for the observed difference and, hence, the name.

Closed form sample size formulas for the test are not available. Required sample sizes must be read from tables (Casagrande et al. 1978; Gail and Gart, 1973; Haseman, 1978) or calculated using computer programs.

Table 9-6 Values of $\Phi(A)$, the proportion of area of a $N(0, 1)$ distribution point lying to the left of a designated point A, for selected values of A

A	$\Phi(A)$	A	$\Phi(A)$
−3.00	0.0013	3.00	0.9987
−2.50	0.0062	2.50	0.9938
−2.00	0.0228	2.00	0.9772
−1.50	0.0668	1.50	0.9332
−1.00	0.1587	1.00	0.8413
−0.75	0.2266	0.75	0.7734
−0.50	0.3085	0.50	0.6915
−0.40	0.3446	0.40	0.6554
−0.30	0.3821	0.30	0.6179
−0.20	0.4207	0.20	0.5793
−0.10	0.4602	0.10	0.5398
0.00	0.5000	0.00	0.5000

9.4.1.2 Chi-square approximation

The standard 2×2 chi-square test (without continuity correction) can be used in place of Fisher's exact text if there are 15 or more patients represented in each of the 4 cells of the table (i.e., there are at least 15 patients in each of the 2 treatment groups who have experienced the event and at least 15 others in each of the 2 treatment groups who have not). This rule is somewhat more stringent than the one proposed by Cochran (1954). He proposed a total sample size of 40 and a cell frequency of ≥ 5. Indications are that, even under these border conditions, the test provides a good approximation to the exact test.

The test can be used for sample size estimation if the event rates in both treatment groups are at or between 0.2 and 0.8 and provided the resulting estimates satisfy the above cell conditions. Fisher's exact test or one of the other tests discussed in this section should be used if the conditions are not satisfied. The formulas for uniform and nonuniform allocation, derived from the 2×2 chi-square test, are as follows:

Uniform allocation ($\lambda = 1$)

$$n_c = \left(Z_\alpha \sqrt{2\bar{P}\bar{Q}} + Z_\beta \sqrt{P_cQ_c + P_tQ_t}\right)^2 / \Delta_A^2 \qquad (9.6)$$

$$n_t = n_c$$
$$N = (r + 1)n_c$$

Nonuniform allocation ($\lambda \neq 1$)

$$n_c = \left(Z_\alpha \sqrt{\bar{P}\bar{Q}(\lambda + 1)/\lambda} + Z_\beta \sqrt{P_cQ_c + P_tQ_t/\lambda}\right)^2 / \Delta_A^2 \qquad (9.7)$$

$$n_t = \lambda n_c$$
$$N = rn_t + n_c$$

where

$\lambda = n_t/n_c$, the ratio of the number of patients assigned to a test-treated group to the number assigned to the control-treated group

n_c = required sample size for the control-treated group

n_t = required sample size for one of the test-treated groups

N = total sample size required in all groups combined

α = type I error probability

β = type II error probability

Z_α = point on the abscissa of a $N(0,1)$ curve (i.e., a normal distribution with mean 0 and variance 1) to the right of which is found $100(\alpha)\%$ of the total area under that curve

Z_β = point on the abscissa of a $N(0,1)$ distribution to the right of which is found $100(\beta)\%$ of the total area under that curve

P_c = assumed event rate (expressed as a proportion) for the outcome of interest in the control-treated group

P_t = assumed event rate (expressed as a proportion) for the outcome of interest in the test-treated group

$Q_c = 1 - P_c$

$Q_t = 1 - P_t$

$\bar{P} = (P_c + \lambda P_t)/(1 + \lambda)$, a weighted average of the 2 event rates

$\bar{Q} = 1 - \bar{P}$

$\Delta_A = P_c - P_t$

9.4.1.3 Inverse sine transform approximation

The inverse sine transform (denoted by $\sin^{-1}$ and expressed in radians) is also used as an approximation to Fisher's exact test (Cochran and Cox, 1957). It has the virtue of providing a good approximation to the exact test over a wider range of P values than is the case with the ordinary 2×2 chi-square test—0.05 to 0.95 compared with 0.2 to 0.8—given the same cell size conditions as specified in Section 9.4.1.2.

Uniform allocation ($\lambda = 1$)

$$n_c = \frac{(Z_\alpha + Z_\beta)^2}{2\left(\sin^{-1}\sqrt{P_c} - \sin^{-1}\sqrt{P_t}\right)^2} \qquad (9.8)$$

$$n_t = n_c$$
$$N = (r+1)n_c$$

Nonuniform allocation ($\lambda \neq 1$)

$$n_c = \frac{(Z_\alpha + Z_\beta)^2 (\lambda + 1)/\lambda}{4\left(\sin^{-1}\sqrt{P_c} - \sin^{-1}\sqrt{P_t}\right)^2} \qquad (9.9)$$

$$n_t = \lambda n_c$$
$$N = rn_t + nc$$

The definitions for P_c, P_t, Z_α, Z_β, and λ are the same as for Equations 9.6 and 9.7.

9.4.1.4 Poisson approximation

The Poisson approximation can be used for comparison of proportions that lie below the lower limit (i.e., 0.05) specified for the inverse sine transform, provided n_cP_c and n_tP_t are both ≥ 10 (Gail, 1974). The same approximation may be used for P values lying above the upper limit (i.e., 0.95) for the transform by using a complementary event (i.e., by using $1 - P_c$ and $1 - P_t$ in place of P_c and P_t in the formula).

Uniform allocation ($\lambda = 1$)

$$n_c = \frac{(Z_\alpha + Z_\beta)^2 (P_c + P_t)}{(P_c - P_t)^2} \qquad (9.10)$$

$$n_t = n_c$$
$$N = (r + 1)n_c$$

Nonuniform allocation ($\lambda \neq 1$)

$$n_c = \frac{(Z_\alpha + Z_\beta)^2 (P_c + P_t/\lambda)}{(P_c - P_t)^2} \qquad (9.11)$$

$$n_t = \lambda n_c$$
$$N = rn_t + n_c$$

9.4.2 Continuous outcome measures

The methods described above may be used for trials involving a continuous outcome measure if the investigator plans to base the primary analysis on a comparison of proportions using a binary categorization of the measure. He should use the methods described in this section if the primary outcome is continuous or near continuous. Conversion of continuous data to binary form for analysis purposes is unwise unless a binary categorization is considered to provide the most relevant treatment of the data. Any categorization reduces the amount of information provided by the data and, if used as a basis for sample size calculations, can be expected to yield an overestimate of the required sample size.

Equations 9.12 and 9.13 are derived using a statistical test for comparison of means observed after a specified period of follow-up. Equations 9.14 and 9.15 are derived using a statistical test for mean change from baseline to some specified period of follow-up. Both sets of equations are based on the normal approximation to the t-statistic. The approximation underestimates sample size if the estimated number of patients per treatment group is <30. Other formulations, such as those discussed by Lachin (1981) and Cochran and Cox (1957), can be used in such cases.

9.4.2.1 *Normal approximation for two independent means*

Uniform allocation ($\lambda = 1$)

$$n_c = \frac{2(Z_\alpha + Z_\beta)^2\,\sigma^2}{(\mu_c - \mu_t)^2} \tag{9.12}$$

$$n_t = n_c$$
$$N = (r + 1)n_c$$

Nonuniform allocation ($\lambda \neq 1$)

$$n_c = \frac{(Z_\alpha + Z_\beta)^2\,\sigma^2\,(\lambda + 1)/\lambda}{(\mu_c - \mu_t)^2} \tag{9.15}$$

$$n_t = \lambda n_c$$
$$N = rn_t + n_c$$

where

μ_c = true mean of the outcome measure for control-treated patients

μ_t = true mean of the outcome measure for test-treated patients

σ^2 = variance of the outcome measure for a single individual (assumed to be the same for all patients in both treatment groups)

and where observed expressions of the outcome measure are assumed to be independent of one another and to be normally distributed. See Section 9.4.1.2 for notation.

9.4.2.2 *Normal approximation for mean changes from baseline*

Uniform allocation ($\lambda = 1$)

$$n_c = \frac{2(Z_\alpha + Z_\beta)^2\sigma_d^2}{(\mu_{dc} - \mu_{dt})^2} \tag{9.14}$$

$$n_t = n_c$$
$$N = (r + 1)n_c$$

Nonuniform allocation ($\lambda \neq 1$)

$$n_c = \frac{(Z_\alpha + Z_\beta)^2\,\sigma_d^2\,(\lambda + 1)/\lambda}{(\mu_{dc} - \mu_{dt})^2} \tag{9.13}$$

$$n_t = \lambda n_c$$
$$N = rn_t + n_c$$

where

$\mu_{dc} = \mu_{1c} - \mu_{0c}$ is the true value of the difference in the outcome measure at follow-up and baseline for the control treatment

$\mu_{dt} = \mu_{1t} - \mu_{0t}$ is the corresponding value for the test treatment

μ_{0c} = true baseline mean (observed just before the initiation of treatment) for the outcome measure for patients assigned to the control treatment

μ_{1c} = true follow-up mean (observed after a specified period of follow-up) for the outcome measure for patients assigned to the control treatment

μ_{0t} and μ_{1t} are the corresponding means for patients assigned to the test treatment

$\sigma_d^2 = 2(1 - \rho)\sigma^2$

σ^2 = variance of the outcome measure on a single individual (assumed to be the same for all patients in both treatment groups) at either baseline or follow-up

ρ = correlation coefficient between baseline and follow-up outcome measures on a single individual

and where the baseline and follow-up measurements made on different patients are assumed to be independent of one another and normally distributed.

9.5 POWER FORMULAS

Sometimes the number of patients available for study is fixed by practical considerations. In these cases it is useful to calculate the power that can be expected with the available sample size.

The power functions for the chi-square approximation and inverse sine transforms are discussed by Lachin (1981). The formulations for the Poisson approximation are based on work by Gail (1974). The power function for Fisher's exact test involves a complicated summation formula that is not practical for routine use.

All of the power formulations given involve use of normal approximations in which

$$\text{Power} = 1 - \beta = 1 - \Phi(A)$$

where

$\Phi(A)$ = proportion of area of a $N(0,1)$ distribution that is to the left of a point A

All other notation is as defined in Section 9.4

9.5.1 Binary outcome measures

9.5.1.1 *Fisher's exact test*

Power estimates must be computed or read from tables of the power functions (Casagrande et al., 1978).

9.5.1.2 *Chi-square approximation*

Uniform allocation ($\lambda = 1$)

$$\text{Power} = 1 - \Phi(A)$$

where

$$A = \frac{Z_\alpha\sqrt{2\bar{P}\bar{Q}/n_c} - |P_c - P_t|}{\sqrt{(P_cQ_c + P_tQ_t)/n_c}} \tag{9.16}$$

Nonuniform allocation ($\lambda \neq 1$)

$$\text{Power} = 1 - \Phi(A)$$

where

$$A = \frac{Z_\alpha\sqrt{\bar{P}\bar{Q}/n_c + \bar{P}\bar{Q}/n_t} - |P_c - P_t|}{\sqrt{P_cQ_c/n_c + P_tQ_t/n_t}} \tag{9.17}$$

9.5.1.3 *Inverse sine transform approximation*

Uniform allocation ($\lambda = 1$)

$$\text{Power} = 1 - \Phi(A)$$

where

$$A = Z_\alpha - \frac{2|\sin^{-1}\sqrt{P_c} - \sin^{-1}\sqrt{P_t}|}{\sqrt{2/n_c}} \tag{9.18}$$

Nonuniform allocation ($\lambda \neq 1$)

$$\text{Power} = 1 - \Phi(A)$$

where

$$A = Z_\alpha - \frac{2|\sin^{-1}\sqrt{P_c} - \sin^{-1}\sqrt{P_t}|}{\sqrt{1/n_c + 1/n_t}} \tag{9.19}$$

9.5.1.4 *Poisson approximation*

Uniform allocation ($\lambda = 1$)

$$\text{Power} = 1 - \Phi(A)$$

where

$$A = Z_\alpha - \frac{|P_c - P_t|}{\sqrt{(P_c + P_t)/n_c}} \tag{9.20}$$

Nonuniform allocation ($\lambda \neq 1$)

$$\text{Power} = 1 - \Phi(A)$$

where

$$A = Z_\alpha - \frac{|P_c - P_t|}{\sqrt{(P_c + P_t/\lambda)/n_c}} \tag{9.21}$$

9.5.2 Continuous outcome measures

9.5.2.1 *Normal approximation for comparison of two independent means*

Uniform allocation ($\lambda = 1$)

$$\text{Power} = 1 - \Phi(A)$$

where

$$A = Z_\alpha - \frac{|\mu_c - \mu_t|}{\sqrt{2\sigma^2/n_c}} \tag{9.22}$$

Nonuniform allocation ($\lambda \neq 1$)

$$\text{Power} = 1 - \Phi(A)$$

where

$$A = Z_\alpha - \frac{|\mu_c - \mu_t|}{\sqrt{(n_t + n_c)\sigma^2/(n_tn_c)}} \tag{9.23}$$

9.5.2.2 *Normal approximation for mean changes from baseline*

Uniform allocation ($\lambda = 1$)

$$\text{Power} = 1 - \Phi(A)$$

where

$$A = Z_\alpha - \frac{|\mu_{dc} - \mu_{dt}|}{\sqrt{2\sigma_d^2/n_c}} \tag{9.24}$$

Nonuniform allocation ($\lambda \neq 1$)

$$\text{Power} = 1 - \Phi(A)$$

where

$$A = Z_\alpha - \frac{|\mu_{dc} - \mu_{dt}|}{\sqrt{(n_t + n_c)\sigma_d^2/(n_tn_c)}} \tag{9.25}$$

9.6 SAMPLE SIZE AND POWER CALCULATION ILLUSTRATIONS

The examples that follow are designed to illustrate sample size and power calculations using the formulas provided in Sections 9.4 and 9.5. Values reported for n_c in illustrations 1 through 5 were rounded up to the next higher integer regardless of the size of the decimal fractions yielded by n_c in the calculations.

9.6.1 Illustration 1: Sample size calculation using chi-square and inverse sine transform approximation

a. Design specifications

- Number of treatment groups (see Section 9.3.1): 2 (i.e., one control and one test treatment)
- Outcome measure (see Section 9.3.2): death
- Follow-up period (see Section 9.3.3): 5 years
- Alternative treatment hypothesis (see Section 9.3.4): one-sided
- Detectable treatment difference in binary outcome (see Section 9.3.5):

$P_c = 0.40$ (5-year control treatment mortality rate)
$\Delta_A = P_c - P_t = 0.10$

- Error protection (see Section 9.3.6): $\alpha = 0.05$, $\beta = 0.05$
- Allocation ratio (see Section 9.3.7): 1:1 (i.e., $\lambda = 1$, equal numbers in test and control groups)
- Losses to follow-up (see Section 9.3.8): 0%
- Losses due to dropouts and noncompliance (see Section 9.3.9): 20%
- Treatment lag time (see Section 9.3.10): 0

b. Method of calculation

Equations 9.6 and 9.8

c. Results

Chi-square approximation (Equation 9.6, Section 9.4.1.2)

$$n_c = \left(1.645\sqrt{2(0.35)(0.65)} + 1.645\sqrt{0.40 \cdot 0.60 + 0.30 \cdot 0.70}\right)^2 / 0.10^2 = 490$$

$n_c = (1/0.8) \times 490 = 613$ (adjusted for 20% loss)
$n_t = 613$
$N = n_c + n_t = 1226$

Inverse sine approximation (Equation 9.8, Section 9.4.1.3)

$$n_c = \frac{(1.645 + 1.645)^2}{2\left(\sin^{-1}\sqrt{0.40} - \sin^{-1}\sqrt{0.30}\right)^2}$$

$n_c = 491$
$n_c = (1/0.8) \times 491 = 614$ (adjusted for 20% loss)
$n_t = 614$
$N = n_c + n_t = 1228$

9.6.2 Illustration 2: Sample size calculation using Poisson approximation

a. Design specifications

Same as for Illustration 1 except:

- Detectable treatment difference (see Section 9.3.5)
  $P_c = 0.04$
  $\Delta_A = P_c - P_t = 0.016$

b. Method of calculation

Equation 9.10, Section 9.4.1.4

c. Results

$$n_c = \frac{(1.645 + 1.645)^2\,(0.040 + 0.024)}{(0.040 - 0.024)^2}$$

$n_c = 2707$
$n_c = (1/0.8) \times 2707 = 3384$ (adjusted for 20% loss)
$N = 3384 + 3384 = 6768$

9.6.3 Illustration 3: Sample size calculation using Coronary Drug Project design specifications

a. Design specifications (Coronary Drug Project Research Group, 1973a)

- Number of treatment groups (see Section 9.3.1): 6 (i.e., 1 control and 5 test treatments)
- Outcome measure (see Section 9.3.2): death
- Follow-up period (see Section 9.3.3): minimum of 5 years
- Alternative treatment hypothesis in binary outcome (see Section 9.3.5): one-sided
- Detectable treatment difference in binary outcome (see Section 9.3.4):
  $P_c = 0.30$ (5-year control treatment mortality rate)
  $\Delta_R = \frac{P_c - P_t}{P_c} = 0.25$
- Error protection (see Section 9.3.6): $\alpha = 0.01$, $\beta = 0.05$
- Allocation ratio (see Section 9.3.7): 1:1:1:1:1:2.5 (i.e., $\lambda = 1/2.5$, for a control group that is 2.5 times as large as any of the five treatment groups)
- Losses to follow-up (see Section 9.3.9): 0%
- Losses due to dropouts and noncompliance (see Section 9.3.9): 30% after 5 years of follow-up
- Treatment lag time (see Section 9.3.10): 0

b. Method of calculation

Equation 9.7, Section 9.4.1.2

c. Results

$$n_c = \left(2.326[0.279 \cdot 0.721 \cdot (0.400+1)/0.400]^{1/2} + 1.645[0.300 \cdot 0.700 + 0.225 \cdot 0.775/0.400]^{1/2}\right)^2 / 0.075^2$$

$n_c = 1906$
$n_t = (1/2.5) \times 1906 = 762$
$n_c = (1/0.7) \times 1906 = 2723$ (adjusted for 30% loss)
$n_t = (1/0.7) \times 762 = 1089$ (adjusted for 30% loss)
$N = 5n_t + n_c = 5(1089) + 2723 = 8168$

The calculations shown above yield results quite similar to those in the Coronary Drug Project using a different method. The total number of patients derived via that method, after adjustment for losses, was $5(1117) + 2793 = 8378$.

9.6.4 Illustration 4: Sample size calculation for blood pressure change

a. Design specifications

- Number of treatment groups (see Section 9.3.1): 2 (i.e., 1 control and 1 test treatment)
- Outcome measure (see Section 9.3.2): blood pressure change after 3 years of treatment
- Follow-up period (see Section 9.3.3): 3 years
- Alternative treatment hypothesis in mean change from baseline (see Section 9.3.4): two-sided
- Detectable treatment difference in continuous outcome measure (see Section 9.3.5):
  $\Delta_A = \mu_{dc} - \mu_{dt} = 4\ mm\ Hg$ (expected difference in mean change from baseline)
  $\sigma^2 = 100\ mm\ Hg^2$ (variance of a single blood-pressure measurement)
  $\rho = 0.3$ (correlation between a baseline blood-pressure measure and the measure after 3 years of follow-up, both taken on the same individual)
  $\sigma_d^2 = 2(1 - \rho)\sigma^2 = 2(0.70)100\ mm\ Hg^2 = 140\ mm\ Hg^2$
- Error protection (see Section 9.3.6): $\alpha = 0.05$, $\beta = 0.05$
- Allocation ratio (see Section 9.3.7): 1:1 (i.e., $\lambda = 1$)
- Losses to follow-up due to dropouts and noncompliance (see Sections 9.3.8 and 9.3.9): 30%
- Treatment lag time (see Section 9.3.10): 0

b. Method of calculation

Equation 9.14, Section 9.4.2.2

c. Results

$$n_c = \frac{2(1.960 + 1.645)^2\,(140\ \text{mm Hg})^2}{(4\ \text{mm Hg})^2}$$

$n_c = 228$
$n_c = (1/0.7) \times 228 = 326$ (adjusted for 30% loss)
$n_t = 326$
$N = 326 + 326 = 652$

9.6.5 Illustration 5: Sample size calculation using Fisher's exact test

a. Design specifications

- Number of treatment groups (see Section 9.3.1): 2 (i.e., 1 control and 1 test treatment)
- Outcome measure (see Section 9.3.2): death
- Follow-up period (see Section 9.3.3): 2 years
- Alternative treatment hypothesis in binary outcome (see Section 9.3.4): one-sided
- Detectable treatment difference in binary outcome (see Section 9.3.5):
  $P_c = 0.5$
  $P_t = 0.1$
  $\Delta_A = 0.4$
- Error protection (see Section 9.3.6): $\alpha = 0.05$, $\beta = 0.10$
- Allocation ratio (see Section 9.3.7): 1:1 (i.e., $\lambda = 1$, equal numbers in the test and control groups)
- Losses to follow-up (see Section 9.3.8): 0%
- Losses due to dropouts and noncompliance (see Section 9.3.9): 0%
- Treatment lag time (see Section 9.3.10):0

b. Method of calculation

Use tables produced by Haseman (1978) or Casagrande and co-workers (1978) and compare the result with that obtained using the chi-square and inverse sine transform approximation.

c. Results

$N = 50$ (25 in each group) from sample size tables in Haseman (1978) or Casagrande et al. (1978)
$N = 42$ (21 in each group) chi-square approximation (Equation 9.6, Section 9.4.1.2)
$N = 40$ (20 in each group) from inverse sine transform approximation (Equation 9.8, Section 9.4.1.3)

Note that $n_t p_t = 4$ is below the limit specified for use with the chi-square and inverse sine transform approximations and that they underestimate the required sample size.

9.6.6 Illustration 6: Power calculation based on chi-square and inverse sine transform approximation

a. Design specifications

- Number of treatment groups (see Section 9.3.1): 2 (i.e., 1 control and 1 test treatment)
- Outcome measure (see Section 9.3.2): death
- Follow-up period (see Section 9.3.3): 5 years
- Alternative treatment hypothesis (see Section 9.3.4): two-sided
- Detectable treatment difference in binary outcome measure (see Section 9.3.5):
  $P_c = 0.40$
  $P_t = 0.30$
  $\Delta_A = P_c - P_t = 0.4 - 0.3 = 0.1$
- Error protection (see Section 9.3.6): $\alpha = 0.05$, β to be determined
- Allocation ratio (see Section 9.3.7): 2:1 (i.e., $\lambda = 2$, twice as many patients in the test-treated group as in the control-treated group), with:
  $n_c = 300$
  $n_t = 600$
  $N = n_t + n_c = 900$
- Losses to follow-up (see Section 9.3.9): 0%
- Losses due to dropouts and noncompliance (see Section 9.3.9): 0%
- Treatment lag time (see Section 9.3.10): 0

b. Method of calculation

Equations 9.17 and 9.19

c. Results

Chi-square approximation:

$$A = [1.960(0.333 \cdot 0.667)(1/300 + 1/600)^{1/2} - |0.400 - 0.300|]/[0.400 \cdot 0.600/300 + 0.300 \cdot 0.700/600]^{1/2}$$
$$= -1.0242$$
$$\text{Power} = 1 - \Phi(-1.0242) = 1 - 0.15 = 0.85$$

Inverse sine transform approximation:

$$A = 1.96 - \frac{2|\sin^{-1}\sqrt{0.40} - \sin^{-1}\sqrt{0.30}|}{\sqrt{1/300 + 1/600}}$$
$$= -1.012$$
$$\text{Power} = 1 - \Phi(-1.012) = 1 - 0.16 = 0.84$$

9.6.7 Illustration 7: Power for design specifications given in Illustration 2 for 1500 patients per treatment group

a. Design specification

As given in Illustration 2 except:

- β to be determined for indicated sample size
- n_c and $n_t = 1500$ (effective sample size, i.e., after reduction for 20% loss due to dropout and noncompliance)

b. Method of calculation

Equation 9.20

c. Results

$$A = 1.645 - |0.040 - 0.024| / [(0.040 + 0.024)/1500]^{1/2}$$
$$= -0.8045$$
$$\text{Power} = 1 - \Phi(-0.8045) = 1 - 0.21 = 0.79$$

9.6.8 Illustration 8: Power for design specifications given in Illustration 4 for 150 patients per treatment group

a. Design specification

As given in illustration 4 except:

- β to be determined for indicated sample size
- n_c and $n_t = 150$ (effective sample size, i.e., after reduction for 30% loss due to dropout and noncompliance)

b. Method of calculation

Equation 9.24

c. Results

$$A = 1.96 - \frac{4 \text{ mm Hg}}{\sqrt{2(140 \text{ mm Hg}^2)/150}}$$
$$= -0.9677$$
$$\text{Power} = 1 - \Phi(-0.9677) = 1 - 0.17 = 0.83$$

9.7 POSTERIOR SAMPLE SIZE AND POWER ASSESSMENTS

The calculations made when the trial is planned will provide the recruitment goal. However, the goal may have to be changed during the trial

For example, it may have to be raised if the observed event rate for the control-treated group during the early stages of recruitment is lower than expected or there is more loss of precision due to noncompliance and dropout than originally envisioned. The period of follow-up may have to be extended as well. Extension of follow-up may be the only option available if recruitment has been completed when the shortfall in desired error protection is first recognized.

There are occasions where an overestimate of P_c in the planning stage may be offset by lower than expected dropout and noncompliance rates during the trial. For example, this was the case in the CDP. The actual five-year mortality in the placebo-treated group was lower than expected, but so were the dropout and noncompliance rates (Coronary Drug Project Research Group, 1973a, 1975).

Power calculations should be made at the end of the trial using the observed sample size and actual losses due to noncompliance and dropouts. Such calculations should be a part of any finished report where the observed treatment effect is small and the authors, therefore, conclude in favor of the null hypothesis of no difference among treatment groups. The calculations, as noted by Freiman and co-workers (1978), are useful to readers when trying to decide whether or not to accept the author's conclusion. A reader may be inclined to accept the conclusion if the estimated power of the study was large enough to detect an important difference, but not otherwise (see also Mosteller et al., 1980).

10. Randomization and the mechanics of treatment masking

Chance favours only those who know how to court her.

Charles N.

10.1 INTRODUCTION

A valid trial requires a method for assignment of patients to a test or control treatment that is free of selection bias. The best method for ensuring bias-free selection is via a bona fide randomization scheme as discussed in Section 8.4 of Chapter 8. Nonrandom methods may be used, but they all suffer from defects that can be avoided with randomization. Hence, randomization is the only method of assignment discussed in this chapter.

Two general designs exist for randomization of patients to treatment: adaptive randomization and fixed randomization. With fixed randomization schemes, the assignment probabilities remain fixed over the course of the trial. In adaptive randomization schemes (also referred to as dynamic randomization, but not in this book) assignment probabilities for the treatments change as a function of the distribution of previous assignments, observed baseline characteristics, or observed outcomes.

The emphasis in this chapter is on fixed randomization. Only a brief overview of adaptive randomization is provided (Section 10.2). Fixed randomization is easier to manage than adaptive randomization. Assignment schedules can be generated before the start of patient recruitment. This is not possible with most adaptive schemes. Assignment must be generated as needed. Further, the generation process is usually complicated enough so that it has to be done on a computer to keep track of previous assignments and any other data used in the adaptation process. All of the trials listed in Appendix B, except one (the National Cooperative Gallstone Study), used fixed allocation schemes. None of the 113 reports of trials reviewed in Chapter 2 gave any indication of having used adaptive randomization. However, this count may be somewhat deceptive in that many of the reports lacked the details needed to reach a definitive statement regarding the method of treatment assignment used.

10.2 ADAPTIVE RANDOMIZATION

There are three general types of adaptive randomization:

- Those in which the assignment probabilities are modified as a function of observed departures from the desired allocation ratio (number adaptive)
- Those in which the assignment probabilities are modified as a function of differences in the observed distribution of baseline characteristics among the treatment groups (baseline adaptive)
- Those in which the assignment probabilities are modified as a function of observed outcomes in the treatment groups (outcome adaptive)

The biased coin randomization procedure, proposed by Efron (1971), is an example of a number adaptive scheme. It is an alternative to blocking in a fixed randomization design (see Section 10.3.3). Patients are assigned to the treatment groups with preset probabilities so long as the difference in the number of patients assigned to the treatment groups remains within a specified range. The probability of assignment to a test treatment is increased or decreased, relative to that for the control treatment, when the range is exceeded.

Baseline adaptive randomization is designed to make certain that the treatment groups are balanced with regard to important baseline characteristics that may affect the outcome measure. In this approach, the assignment probabilities are a function of observed differences in the baseline composition of patients already enrolled (Begg and Iglewicz, 1980; Freedman and White, 1976; Friedman et al., 1982; Pocock, 1983; Pocock and Simon, 1975; Simon, 1977). The main advantage of the technique is the opportunity it provides for balancing the composition of treatment groups on several different baseline characteristics without stratification (see Section 10.3.2). The main disadvantage is in its administrative complexities. The technique cannot be managed without a computer.

The play-the-winner scheme, proposed by Zelen (1969), is an example of outcome adaptive randomization. The simplest version is one involving only one test and one control treatment, where the first patient enrolled has the same probability of being assigned to either treatment, and thereafter the assignment received by each patient is a function of the outcome observed and the treatment assignment of the preceding patient. The assignment will be the same as for the preceding patient if the outcome observed for that patient was favorable. The assignment will be to the other treatment if the outcome was unfavorable. Hence, the name, play-the-winner.

The main difficulty with the scheme, at least with simple versions such as the one described, is that it allows an investigator to predict the next assignment, thereby introducing the possibility of bias into the patient selection process. A second limitation is the need to determine the outcome for the last patient enrolled before the next one can be enrolled.

The play-the-winner algorithm has been modified to incorporate outcome information from multiple patients (Wei and Durham, 1978). This modification eliminates dependence on the last outcome observed and therefore makes it more difficult for an investigator to predict the next assignment. However, even modified in this way, the scheme has limited utility. The ability to identify a "winning" treatment and to have that knowledge influence treatment assignments during the patient recruitment process is minimal in most trials requiring long-term follow-up for outcome assessment.

10.3 FIXED RANDOMIZATION

Fixed randomization schemes require specification of the:

- Allocation ratio
- Allocation strata
- Block size

The considerations involved in making these specifications are outlined in the subsections that follow.

10.3.1 Allocation ratio

The number of allocations made to any one of the study treatments is a function of the assignment probabilities—assumed to be set in advance of patient recruitment and to be held fixed over the course of recruitment in fixed allocation schemes. The only changes that occur are due to major design modifications, such as occurred in the University Group Diabetes Program (UGDP) with the addition of a fifth treatment (phenformin) some 18 months after the start of patient recruitment (University Group Diabetes Program, 1970d).

The allocation of patients to the study treatments can be uniform or nonuniform. A design will be characterized as uniform if the assignment probabilities for the t test treatments and control treatment are equal, i.e.,

$$P_1 = P_2 = \cdots = P_i = \cdots = P_{t+1} \qquad (10.1)$$

where

P_i, $i = 1, \cdots, t$, denote assignment probabilities for the t test treatments

and

P_{t+1} denotes the assignment probability for the control treatment

and where

$$\sum_{i=1}^{t+1} P_i = 1$$

It will be characterized as nonuniform if there is at least one probability value in Equation 10.1 that differs from the other values in the equation.

The entire allocation scheme for the trial can be expressed as a ratio of $t + 1$ numbers,

$$r_1 : r_2 : \cdots : r_i : \cdots : r_{t+1}$$

where r_i is the expected number of assignments to the ith test treatment and where all values of r are expressed as integers, reduced so as to have no multiplier in common other than 1 (e.g., an allocation ratio involving 1 assignment to the test treatment for every 2 assignments to the control treatment would be expressed as a ratio of 1:2). Expressed this way,

$$\sum_{i=1}^{t+1} r_i = B \qquad (10.2)$$

where B is the minimum block size (see Glossary and Section 10.3.3). For example, the minimum block size in a 2-treatment trial with an allocation ratio of 1:1 is 2. It is 4 if the allocation ratio is 1:3. It is 5 for a 3-treatment trial with an allocation ratio of 2:2:1.

All $t + 1$ values of r are equal to 1 in uniform fixed allocation designs. At least one value of r will be greater than 1 in nonuniform fixed allocation designs.

The most common allocation design is one involving uniform allocation. All of the trials sketched in Appendix B, except three, were of this type (see line 20e, Table B-4, Appendix B). Uniform allocation should be used, except where there are valid reasons to allocate a disproportionately larger number of patients to one treatment than to another. The reasons may have to do with the cost of one treatment versus another, the way of administering one versus another, or the presumed safety or efficacy of one versus another (see Persantine Aspirin Reinfarction Trial Research Group, 1980a, for example of nonuniform allocation). Other reasons relate to statistical considerations, as discussed in Section 9.3.12, where the study involves multiple test treatments, each of which is to be contrasted with the same control treatment. A third set of reasons relate to secondary research aims that are best pursued via use of nonuniform allocation. One of the reasons why the Coronary Drug Project (CDP) enrolled more patients in the placebo-treated group than in any of the test-treated groups had to do with a secondary aim (Coronary Drug Project Research Group, 1973a).

10.3.2 Stratification

Stratification[1] during patient enrollment involves the placement of patients into defined strata for randomization. It is done to reduce or eliminate variation in the outcome measure due to the stratification variable(s) (see Table 10-1 for points concerning stratification during randomization). A variable is said to be controlled when patients are assigned to treatment in such a way so as to ensure that it has the same distribution in all treatment groups. Separate allocation schedules are required for the various levels or states assumed by the variables to be controlled. Allocations to each stratum are made using the same allocation ratio as for all other strata. A scheme requiring control of sex would require a separate allocation stratum for males and for females. A scheme requiring control of sex and age, the latter classified at three levels (e.g., <45, 45 through 55, and >55), would require six (i.e., 2·3) allocation strata, one for each age level and sex combination. In general terms, s stratification variables with l_i levels for the ith variable will produce a total of $l_1 \cdot l_2 \cdots l_i \cdots l_{s-1} \cdot l_s$ allocation strata.

The term *stratification*, as used throughout this chapter, refers to a process that takes place in conjunction with randomization, and that is based on data collected prior to randomization. Stratification that is done in conjunction with data analyses, as discussed in Chapter 18, is referred to as post-stratification. Both forms of stratification may be used in the same trial, but not on the same variable.

The main arguments for stratification involve a combination of philosophic and statistical considerations. Ideally, the goal in any trial is to carry out the comparison of the study treatments in groups of patients that are identical with regard to all entry characteristics that influence the outcome measure. The best way to achieve the goal is via matching for all variables of concern. However, it is impractical for reasons discussed in Chapter 8. The best that can be done is to stratify the study groups on a few variables and then to randomize within those strata.

Clearly, there is a practical limit to the number of variables that can be realistically controlled via stratification. The number of strata quickly reaches unmanageable limits when a number of different variables are used. As a result, the choice of variables must be judicious and by definition must be limited to variables that are independent of the treatment assignment. In addition, the choice should be limited to variables that are not subject to large observational or recording errors so as to minimize clas-

[1] Not to be confused with poststratification (see Glossary).

Table 10-1 Stratification considerations for randomization

- Only variables that are observed and recorded before randomization may be used for stratification in the treatment assignment process.
- Increased statistical efficiency resulting from stratification is minimal for trials involving ≥50 patients per treatment group.
- It is impractical to control for more than a few sources of variation via stratification at the time of randomization (i.e., generally no more than two or three).
- Use of a large number of allocation strata may allow for fairly large chance departures from the desired allocation ratio if there are only a small number of patients per stratum.
- Any gain in statistical efficiency resulting from stratification using a given variable will be a function of the relationship of that variable to the outcome measure. The gain will be small to nil if the relationship is weak or nonexistent. It will be greatest for variables that are highly predictive of outcome.
- Stratification on any patient characteristic complicates the randomization process; it may prolong the time needed to clear a patient for enrollment if stratification depends on readings or determinations made outside the clinic.
- Variables used for stratification should be easy to observe and reasonably free of measurement error.
- Variables that are subject to major sources of error due to differing interpretations should not be used for stratification. They are of limited use for variance control and the errors made may open the study to criticism when the results are published.
- It is unreasonable to expect that all important sources of baseline variation can be controlled via stratification during randomization. Analysis procedures involving post-stratification and multiple regression will be required to adjust treatment comparisons for baseline differences not controlled via stratification.
- Use of any stratification scheme that involves calculations or complicated interpretations should be avoided, especially in self-administered randomization schemes where the calculations or interpretations are not checked before treatment assignments are issued.
- Clinic should be used for stratification in multicenter trials. This form of stratification will control for differences in the study population due to environmental, social, demographic, and other factors related to clinic.

sification errors made in the stratification process.

The gain in statistical precision from stratification is inconsequential once the number of patients per treatment group reaches 50 or more. The greatest gains are for small trials involving 20 or fewer patients per treatment group (Grizzle, 1982; Meier, 1981).

Clinical trial researchers are divided over the wisdom of stratification at the time of randomization. Those in favor of the process presume that even if it does not increase statistical precision it is unlikely to reduce it. Therefore, why not stratify? Those who question use of the process argue that the statistical gain, at best, is likely to be small. This fact, coupled with the practical complexities involved in administering the process, serve as the main arguments against stratified randomization (see Brown, 1980, for pro arguments; Meier, 1981, and Peto and coworkers, 1976, for con arguments). The diversity of opinion is reflected in the trials sketched in Appendix B. Six of the trials did not stratify on any patient characteristic. The other eight used sex, age at entry, and/or some indicator of disease state for stratification (see item 20.b, Table B-4, Appendix B).

The goal in stratification is to reduce the variance associated with treatment comparisons through control of variables that affect outcome. Clearly, there will be no reduction, and hence no gain in statistical precision, if the variables are unrelated to outcome. The more restrictive the patient selection criteria, the less the need for any stratification. The relationship of a variable (e.g., age) to an outcome (e.g., death), even if quite striking when assessed over a broad range of unselected patients, may be modest over the range represented by patients enrolled into the trial.

The CDP provides graphic evidence of the futility of identifying factors that predict mortality, the outcome of interest in that study and several of the others sketched in Appendix B. A multiple linear regression model, using 40 different baseline characteristics as predictors for mortality, accounted for only 10.6% of the observed variance associated with mortality (Coronary Drug Project Research Group, 1974). Risk group, defined by number and severity of previous myocardial infarctions and the only variable used for stratification other than clinic, had little predictive value. It ranked 26 in the list of 40 variables in terms of predictive value. The five most important predictors, in order of importance using a stepwise regression procedure were: ECG ST segment depression, cardiomegaly (as read from chest x-rays), New York Heart Association functional class, ventricular conduction defects (as read from ECGs), and history of use of diuretics. They accounted for over two-thirds of the total variance explained by the model.

Stratification using patient characteristics should not be undertaken lightly. It will complicate the randomization process since assignments cannot be made until all data needed for stratification are in hand. This may delay, sometimes by weeks, the enrollment of a patient if needed data come from laboratory determinations or readings made outside the clinic. Variables that require a series of complex and error prone classifications in order to be converted into values suitable for use in stratification should be avoided. The same is true for variables requiring subjective interpretations. A high error rate in the classification of patients by strata can negate the effect of stratification and may open the study to criticism when the results are published.

Clinic is a natural stratification variable in multicenter trials. All of the 13 multicenter trials sketched in Appendix B (see item 20.b, Table B-4) used this form of stratification. The cautions expressed above with regard to use of patient characteristics for stratification do not apply to clinic. Use of separate allocation schedules by clinic, with each schedule having the same allocation ratio, ensures comparability of the treatment groups with regard to the mix of patients coming from the various clinics in the trial. This assurance is important since clinic populations can differ widely with regard to a host of characteristics, even if the study has fairly rigid entry criteria. Patients will come from different geographic areas and, hence, will have different environmental exposures and perhaps demographic characteristics as well. Further, there may be subtle differences in treatment patterns from clinic to clinic, even if the study has a well defined treatment plan. In addition, there are practical reasons for the stratification, especially in masked drug trials in which clinics receive the drugs they are to use in coded bottles from a central supply point. It is much easier for the supplier to estimate the drug needs of individual clinics if the allocation ratio is fixed across clinics than when it is not.

Clinic variation in outcome event rates can be seen from inspection of the UGDP results. The number of deaths recorded ranged from a high of 23 out of the 90 patients enrolled in the Cincinnati clinic to a low of 1 out of the 87 patients enrolled in the Baltimore clinic when the first results from the trial were published (University Group Diabetes Program Research Group, 1970e). Four of the 12 clinics accounted for a little over 70% of all deaths reported. Critics of the study cited clinic variation in mortality as one of the explanations for the tolbutamide results (see Chapter 7). However, in doing so they failed to recognize that the variation was unlikely to be treatment-related because of the stratification by clinic in the randomization process.

Normally, the question of who treats within a clinic is ignored in the randomization process. None of the trials sketched in Appendix B controlled for this source of variation. Physician-to-physician variation in treatment practice may be small in masked drug trials, but may not be in unmasked trials, especially those involving surgical procedures. It may be appropriate in such cases to control for anticipated variation by stratifying on treating physician.

Statistical considerations are only one reason for stratification. It is sometimes done simply as a ploy to protect the study from criticism when it is finished. Indeed, it is easier to answer criticisms concerning the comparability of the study groups if the criticisms focus on variables that have been stratified. However, defensive stratification can backfire if the variables selected are viewed by critics as "inappropriate" or if they are able to make cogent arguments suggesting that other "more important" variables were left uncontrolled.

Stratification is also used to control for a variable known or suspected to interact with treatment (see Glossary for definition of treatment interaction). Stratification of this sort should be considered for any variable that, depending on its level, has the potential of ameliorating or enhancing a treatment effect. The experimenter, via stratification, is able to compare treatment effects across strata and thereby estimate the size of the interaction effect. In actual fact, however, most interactions, unless they are pronounced, are difficult to detect. The typical trial, because of its small size, provides little statistical power for their detection.

Extreme cases of interactions in which the treatment has a positive effect when the interacting variable assumes one state and the opposite effect when it assumes another state should not be controlled via stratification. They should be dealt with by constructing more restrictive selection criteria so only patients who react positively to the treatment are enrolled.

10.3.3 Block size

The investigator must decide whether to constrain the randomization process so as to ensure balance in the number of allocations made to the various treatment groups in a stratum at various points over the course of patient enrollment. Unconstrained randomization may lead to imbalances in the baseline characteristics of the treatment groups if there are, quite by chance, long unbroken runs of assignments to the same treatment and if the type of patients enrolled changes over time. Table 10-2 lists considerations involved in blocking.

The desired allocation ratio in a stratum could be achieved with a single blocking constraint if the exact number of patients to be enrolled in the stratum were known in advance. However, this approach is not recommended. First, there are few situations in which it is practical to recruit to a set limit within a stratum. Hence, failure to achieve the desired recruitment goal could mean that the study closes far from the desired allocation ratio. Second, the approach may allow too much room for variation around

Table 10-2 Blocking considerations

Blocking should be considered if:

- Patient enrollment is likely to continue over an extended period of time, or if the demographic or clinical characteristics of the study population can be expected to change over the course of enrollment
- There are practical or statistical reasons why it is important to satisfy the specified allocation ratio at various points during the enrollment process

Block size considerations:

- The smallest possible block size is the sum of integers defined by the allocation ratio (see Equation 10.2)
- The block sizes used for construction of an allocation schedule should not be divulged until it is appropriate to do so—and never before patient enrollment is completed
- The larger the block, the greater the chance of departure from the specified allocation ratio
- Variable block sizes are preferable to fixed blocks, especially in unmasked trials
- Use of a large number of allocation strata may lead to a large departure from the specified allocation ratio, unless small block sizes are used within each stratum

the desired allocation ratio over the course of patient enrollment. For example, the constraint in a trial involving two treatments, a 1:1 allocation ratio, and a single block of 100 patients does not take effect until 50 assignments have been made to one of the two treatment groups. Hence, in theory it is possible that the results of the trial could be completely confounded with time of enrollment if the first 50 patients are assigned to the same treatment. A third reason has to do with the need for interim analyses over the course of the trial, as discussed in Chapter 20. These analyses are easier to interpret if large departures from the desired allocation ratio have been avoided. Certainly, blocking is recommended any time recruitment extends over a long period of time.

The usual approach to blocking in fixed allocation schemes is to use a sequence of blocks of the same size or of differing sizes, each of which is constructed using the same allocation ratio. All of the 14 trials sketched in Appendix B, except two—the National Cooperative Gallstone Study (NCGS) and the Veterans Administration Cooperative Studies Program Number 43 (VACSP No. 43)—used this approach.

The blocking arrangement used should not be revealed to clinic personnel until it is appropriate to do so (after patient recruitment is completed in unmasked trials and after the trial is completed in double-masked trials). Further, the scheme used should be designed to minimize the chance of clinic personnel discovering the blocking scheme. Discovery of the scheme can lead to selection biases if the information is used to predict future assignments and if the predictions influence decisions on enrollment. The probability of making correct predictions is highest with simple blocking schemes involving small blocks of uniform size. For example, it is 0.5 in designs involving two treatment groups and an unmasked treatment assignment scheme using blocks of size two. The chance of discovering the blocking pattern is minimal with large blocks, even if blocks of uniform size are used, especially if treatments are administered in double-masked fashion as in the CDP (see Section 10.5 and Coronary Drug Project Research Group, 1973a).

The preferred approach, particularly in unmasked trials, involves a mix of different block sizes with the order specified. One arrangement is to have the blocks filled in order according to size. This arrangement may be considered if blocks of several different sizes are used and if the largest block represents a sizable fraction of the total numbers of assignments anticipated in a stratum. The arrangement reduces the amount of variation around the specified allocation ratio as recruitment proceeds—a desirable feature if the designers wish to have an observed allocation ratio that is near the specified one when recruitment is finished. An alternative approach involves a random order of blocks according to size. It is preferred to the ordering described above when only two or three different block sizes are used and when each stratum contains several blocks of each size. The random ordering eliminates any chance of clinic personnel discovering the blocking pattern.

The usefulness of blocking can be reduced by the use of too many allocation strata. There can be large departures from the desired allocation ratio if none of the blocks in the individual strata are filled by the time patient recruitment is completed. Use of small block sizes will help guard against this problem, but their use may increase the chances of predicting future assignments, as discussed above.

10.4. CONSTRUCTION OF THE RANDOMIZATION SCHEDULE

The randomization schedule can be constructed once the design specifications outlined in Section 10.3 have been set. Construction may be done using output from:

- A published list of random numbers, e.g., as provided by The Rand Corporation (1955)
- Published random permutations of a set of numbers, e.g., those appearing in Cochran and Cox (1957) and Fisher and Yates (1963)
- A computer-based pseudo-random number generator

Methods such as coin flipping, where the order of assignment cannot be replicated, are unacceptable (see Chapter 8).

Most computer statistical packages include pseudo-random number[2] generators. They may be used for construction of the allocation schedule, but with some caution. Output from some of the generators involves serial correlations (e.g., see Hauck, 1982). While the defect is not of great concern in the allocation process, it is best to use a generator that has been tested for the defect and found to be free of it.

An algorithm is needed to translate output obtained from the randomizing device into treatment assignments. The translation is straightforward for schemes based on tables of random permutations, as in Illustration 1 in Section 10.5.1. It is more complicated for schemes using output from tables of random numbers or from pseudo-random number generators. The method described in Table 10-3 is based on an algorithm proposed by Moses and Oakford (1963) and can be implemented using the worksheet displayed in Table 10-4. Use of the algorithm is illustrated in Table 10-5 for a random sequence of numbers selected from Table 10-6.

2. So termed because the numbers they generate are not the result of a random process, but have properties similar to those generated via a random process.

10.5 MECHANICS OF MASKING TREATMENT ASSIGNMENTS

Masked administration of treatment (see Chapter 8 for discussion of the rationale for masking) is feasible only in cases in which it is possible to administer all study treatments in an identical fashion and in which clinic personnel do not need to know the identity of the treatment being administered in order to care for the patient receiving it. Most applications of masked treat-

Table 10-3 Moses-Oakford assignment algorithm for block of size k

Step	*Illustration (see Table 10-5)*
1 Specify number of treatment groups, $t + 1$.	$t + 1 = 4$
2 Specify treatment allocation ratio, $r_1{:}r_2{:} \cdots {:}r_t{:} \cdots r_{t+1}$ such that $\sum_{i=1}^{t+1} r_i = B$ (see Equation 10.2).	$r_1 = 1$ $r_2 = 1$ $r_3 = 1$ $r_4 = 1$ $B = 4$
3 Specify block size k such that it is $\geq B$ and is divisable by B.	$k = 8$
4 Specify treatment symbols or codes.	C = Control T1 = Test treatment 1 T2 = Test treatment 2 T3 = Test treatment 3
5 Set down an arbitrary sequence of treatment symbols in column 2 of worksheet (Table 10-4), such that the allocation ratio specified in step 2 is satisfied.	
6 Generate a random number,* N_1, such that it is $\geqslant 1$ but $\leqslant k$; record value in column 5, line k, of worksheet.	$N_1 = 1$, record on line 8, col. 5
7 Take treatment symbol on line N_1, column 2, and record on line k, column 4.	C, from line 1, col. 2, record on line 8, col. 4
8 Cross out symbol on line N_1, column 2. Record symbol given on line k, column 2, on line N_1, column 3 (skip if $N_1 = k$).	Cross out C, line 1, col. 2, add T3 to line 1, col. 3
9 Generate a new random number N_2 such that it is ≥ 1 but $\leq k - 1$ and record in column 5, line $k - 1$.	$N_2 = 4$, record on line 7, col. 5
10 Take treatment symbol on line N_2, column 2 or from column 3, if any appear in column 3, record on line $k - 1$, column 4.	T1 from line 4, col. 2, record on line 7, col. 4
11 Cross out the symbol appearing in columns 2 or 3, line N_2. Record symbol given on line $k - 1$, columns 2 or 3 on line N_2, column 3 (skip if $N_2 = k - 1$).	Cross out T1, line 4, col. 2, add T3 to line 4, col. 3
12 Repeat steps 9, 10, and 11 reducing the upper limit of permissible random numbers by 1 for each repetition** until all but the last assignment has been made.	As outlined above
13 Complete the scheme by recording in column 4 the unused treatment symbol appearing on line 1, columns 2 or 3.	Take T3 from line 1, col. 3, record on line 1, col. 4

*Numbers are drawn from page 17 of The Rand Corporation's 1 million random digits (1955), as reproduced in Table 10-6.

**The algorithm is written to allow the user to work from the bottom up on the worksheet illustrated in Table 10-4. It can be written to allow use of the sheet from the top down but this arrangement complicates keeping track of the permissible range for the next random number to be selected. With the methods as outlined, the limit for the next number to be selected is given by the line number of the next line on the sheet to be filled.

Table 10-4 Moses-Oakford treatment assignment worksheet for block of size *k*

Block size	Treatment codes		Random numbers*		
______			Page	Column	Row
Allocation ratio		Start	___	___	___
______		End	___	___	___
		Source:	______		

(1)	(2)	(3)	(4)	(5)
	Treatment assignments			
Order of assignment	Initial	Replacements	Final	Random number
1	___	______	___	___
2	___	______	___	___
3	___	______	___	___
4	___	______	___	___
5	___	______	___	___
6	___	______	___	___
7	___	______	___	___
8	___	______	___	___
⋮	⋮	⋮	⋮	⋮
k-2	___	______	___	___
k-1	___	______	___	___
k	___	______	___	___

*Reading rule:

ment administration arise in the context of drug trials. Masking is accomplished by bottling, packaging, labeling, and dispensing the test and control drug in an identical fashion. Tablets may have to be formulated using a taste-masking substance, such as quassin as in the CDP Aspirin Study (Coronary Drug Project Research Group, 1976), to obscure telltale tastes. Another alternative is to use an enteric coating on the tablets, provided the coating does not reduce the bioavailability of the drug. Generally, masking the identity of a drug is easier to accomplish if the drug is contained in capsules than if it is contained in tablets. The capsules help to obscure taste differences that may be present when tablets are used.

There can be subtle differences in sheen, color, or texture of tablets as well. For example, there was a slight difference in the sheen of tolbutamide tablets as contrasted with the corresponding placebo tablets in the UGDP. However, the difference was apparent only in indirect light, and then only in side-by-side comparisons of the two kinds of tablets. Such differences are avoided with opaque capsules.

Trials involving multiple test treatments should be designed with the goal of using a single placebo unless it is not possible or practical to do so. The goal cannot be achieved if the study medications are dispensed in different forms, as in the case of the UGDP. Two kinds of placebo pills were required, one to match tolbutamide tablets and the other to match phenformin capsules.

Use of a common placebo imposes the same pill schedule on all patients, regardless of treatment assignment. For example, the CDP required all patients to take nine capsules per day in order to deliver the required dosage of nicotinic acid. Several of the medications could have been delivered via a smaller number of capsules (Coronary Drug Project Research Group, 1973a). However, this would have required a different placebo for those drugs.

The way in which medications are bottled and labeled is important. There is no value in going to great lengths to develop matching tablets or capsules if the test and control medications arrive at the clinic in different sized or colored bottles. The differences do not have to be great to destroy the mask. Subtle variations in the way the bottles are capped or labeled may be enough to do the job. The best approach is to have all medications bottled and labeled at the same facility, under tightly controlled conditions. The VA Cooperative Studies Program has established a central pharmacy to supply its trials with needed study medications (Hagans, 1974). Various other trials, such as the CDP, have contracted with a single facility to supply drugs to the study clinics (Coronary Drug Project Research Group, 1973a).

In a typical drug trial, clinics will dispense drugs by bottle number. The treatment assignment issued by the data center will indicate the bottle number to be used. The simplest bottle numbering scheme is one in which all bottles containing a given drug bear the same number or letter designation. The trouble with such schemes is that all patients on a drug are unmasked as soon as any one patient on the drug is unmasked. Use of a unique bottle number for every patient in a clinic avoids this problem, but such schemes complicate the logistics of supplying clinics with needed drugs. A compromise between these two extremes was used in the CDP. Each clinic was supplied with sets of bottles, labeled from 1 through 30, as discussed in Illustration 7 of Section 10.8.7. This meant that clinics had somewhere between 5 and 8 patients on the same bottle number by the time recruitment was finished.

Table 10-5 Illustration of Moses-Oakford algorithm

Block size	Treatment codes		Random numbers*		
______	C = Control T1 = Test trt 1 T2 = Test trt 2 T3 = Test trt 3		Page	Column	Row
Allocation ratio		Start	17	27	16
1:1:1:1		End	17	27	24
		Source:	Rand Corp. (1955)		

(1)	(2)	(3)	(4)	(5)
	Treatment assignments			
Order of assignment within block	Initial	Replacements	Final	Random number
1	~~C~~	~~T3~~, T3	T3	___
2	~~C~~	~~T2~~, T2	T2	2
3	~~T1~~	T3	T3	1
4	~~T1~~	T3	T1	3
5	T2	______	T2	2
6	T2	______	C	2
7	T3	______	T1	4
8	T3	______	C	1

* Reading rule:
Read down to the end of column and from left to right. Ignore 0's and numbers in excess of number of lines remaining to be filled.

Table 10–6 First 25 lines of page 17 of The Rand Corporation's 1 million random digits

Row number	Column number 5	10	15	20	25	30	35	40	45	50
	00397	56753	53158	71872	68153	09298	20961	49656	33407	95683
	14328	44708	72952	27048	67887	28741	46752	88177	95894	40086
	88534	87112	68614	83073	88794	96799	67588	75049	84603	83140
	97347	87316	73087	77135	71883	98643	03808	08848	14133	60447
5	01366	72976	01868	51667	63279	60040	88264	79152	03474	61366
	20523	21584	93712	83654	89761	90154	96345	37539	32556	74254
	70603	97122	44978	78028	08943	13778	11080	34271	68266	85372
	48410	94516	15427	75323	71685	70774	50342	33771	03678	42321
	69788	41758	55004	30992	17402	63523	42328	87171	24751	15084
10	33884	83655	88345	69602	52606	57886	18034	03381	75796	35901
	77480	28683	68324	66035	07223	14926	16128	13645	90370	31949
	11057	98849	29499	21565	30786	83292	92392	37104	36899	49906
	79368	43710	80365	88735	75275	21664	57965	19002	00301	12658
	94385	01717	96191	50404	80166	93965	24688	27839	10812	31715
15	92127	42588	93307	80834	11317	26583	25769	98227	14887	58462
	29148	68662	26872	72927	79021	51622	29521	33355	45701	45996
	33782	93424	16530	96086	17329	74020	11501	46660	05583	22277
	77653	55430	84644	00448	86828	58855	67451	95264	67386	82424
	52611	60012	88620	72894	94716	22262	99813	69592	63464	33163
20	91857	47904	22209	78590	68615	52952	31441	41313	18550	72685
	68825	04795	53971	14592	39634	23682	76630	02731	81481	86542
	23727	54291	56045	61635	32186	90355	73416	63532	24340	18886
	84832	30654	48543	18339	65024	91197	64624	74648	09660	27897
	49771	11123	08732	49393	12911	72416	17834	18878	62754	85072
25	23727	56577	51257	83291	12329	16203	91681	68138	79959	43609

Source: Reference citation 387. Reprinted with permission of The Rand Corporation (New York: The Free Press, 1955). Copyright © 1955 and 1983 by The Rand Corporation.

However, it also meant that they could get by with a much smaller inventory of drugs than would have been required with individually numbered bottles.

Most prepackaged medications in masked trials will be supplied to clinics with a two-part label, as illustrated in Figure 10-1. One part of the label will be affixed to the package and dispensed with it. It should bear the name of the study, the bottle or container number, instructions for taking the medication, and the name of the physician or clinic responsible for dispensing the medication. The other part of the label is loosely affixed to the container. Its prime purpose is to indicate container contents, either on the face of the label (for single-masked trials), or by breaking a seal (for double-masked trials). It is required for interstate shipment of drugs under federal law; it is illegal to ship drugs across state lines without it. It is detached when the medication is dispensed and is ordinarily retained at the clinic to allow clinic personnel to unmask a medication in an emergency.

10.6 DOCUMENTATION OF THE RANDOMIZATION SCHEME

There should be a written description of the scheme used to generate the allocation schedule. It should be written when the randomization schedule is produced and should be checked for clarity and accuracy before it is filed for future reference. Table 10–7 provides an outline of the items to be covered in the writeup. The details should be sufficient to allow a person from outside the study to reproduce the schedule with the information provided.

The documentation may be needed to defend the study years after the completion of randomization. The UGDP serves as a case in point. The Committee for the Assessment of Biometric Aspects of Controlled Trials of Hypoglycemic Agents (1975), appointed to review the study about 8 years after the completion of patient enrollment, was especially interested in the randomization process used.

A. Attached portion of bottle label

The XYZ Trial

Bottle number 42

R_x: Take one capsule each morning

For: *Harry L. Green*

Date: *3-7-85* John Smith, M.D.
Phone: 555-1701

B. Detachable portion of bottle label

The XYZ Trial

Bottle number 42

In case of emergency open this label for bottle contents

Call: (301) 955-8889
XYZ Coordinating Center

Figure 10-1 Stylized bottle label for medications dispensed in the XYZ trial.

10.7 ADMINISTRATION OF THE RANDOMIZATION PROCESS

An allocation scheme, no matter how carefully constructed, will be useless as a means of protecting against patient selection bias if it is not followed. Departures from the schedule to accommodate the desire of a patient or his physician, no matter how well motivated, are never justified. They can invalidate the results of the entire trial if they are numerous and if there are reasons to believe they are treatment related. A carefully executed trial will include various safeguards to make certain the assignment schedule is followed, as listed in Table 10–8.

The preferred system is one in which allocations are issued from a central point on a per-patient basis. The main advantage with such systems, as opposed to systems with no central control (e.g., as in systems with envelopes placed in the clinic to be used in the order provided), lies in the audit trail provided and the opportunity to proscribe release of an assignment until a patient has been shown to be eligible for enrollment via the data provided, the required baseline data have been collected, and his consent to participate has been obtained. The CDP used a

Table 10–7 Items that should be included in the written documentation of the allocation scheme

A. For procedures using published lists of random numbers

- Reference citation to the published numbers
- Section of the table or list used (indicate enough detail to allow regeneration of the schedule)
- Reading instructions indicating the order in which numbers are read, including a description of any modular arithmetic used to convert numbers outside the usable range to usable values
- Specifications of the construction process, such as those listed for illustrations in Section 10.8
- Worksheets or computer program used to generate the assignment list
- Copy of the assignment list

B. For procedures using computer based pseudo-random number generators

- Reference citation to the pseudo-random number generator
- Program listing of the pseudo-random number generator
- Seed used to start the generation process
- First and last numbers generated with the seed
- Specifications for the construction process, such as those listed for illustrations in Section 10.8
- Computer programs used to generate the assignment list
- Copy of the assignment list

Table 10–8 Safeguards for administration of treatment allocation schedules

- Avoid the use of any assignment scheme that has a high degree of predictability (e.g., use of small blocks as discussed in Section 10.3.3)
- Keep each treatment assignment masked to the patient, physician, and person issuing the assignment until the patient has been accepted into the study and is ready to start treatment
- Vest responsibility for issuing assignments in an individual or group located outside the clinic
- Withhold disclosure of an assignment until the patient is judged eligible for enrollment, has given his consent to be enrolled, and all essential baseline data have been obtained
- Make certain that the assignment process establishes a clear audit trail that indicates who requested the assignment and when it was issued

MP-100
1414b

centrally administered mail-based assignment scheme (Coronary Drug Project Research Group, 1973a). The Coronary Artery Surgery Study (CASS) used a centrally administered telephone-based assignment scheme (Coronary Artery Surgery Study Research Group, 1981). Either scheme is preferable to one that is self-administered. Such systems are subject to the abuses noted in Section 8.4.

Table 10-9 contains a facsimile of an allocation schedule from the CDP, as used in the Coordinating Center for making assignments. The allocation process required the clinic to initiate the request. This was done by sending the forms completed for a patient's two prerandomization visits to the Coordinating Center. An allocation was not released by the Center if essential items of information were missing from the forms, if an eligibility stop condition (see Section 12.5.8) had been checked, or if the clinic did not indicate that a signed consent had been obtained from the patient indicating his willingness to be enrolled into the trial.

Once all essential conditions were met, a treatment assignment form was prepared (Part A, Table 10-10). The bottle assignment recorded on the form was taken from the first topmost empty line of the allocation schedule for the clinic and stratum to which the patient belonged (the third line in the sample schedule in Table 10-9). The ID number and the name of the patient were entered on the line. After entry of the required data on the treatment assignment form, it was placed in an opaque envelope (Part B, Table 10-10), which was then sealed and placed in a larger envelope for mailing to the clinic. The inner envelope was retained in sealed condition at the clinic until the patient returned for his final baseline examination and was judged ready to start treatment. A patient was not considered enrolled in the trial until the clinic opened the treatment allocation envelope. Once this was done, the patient was counted as a member of the treatment group to which he had been assigned. Assignments issued for patients who failed to return for their last baseline visit, or who withdrew their consent at that visit, were not counted, provided they were returned to the Coordinating Center in sealed condition. The ID numbers and names of such patients were deleted from the allocation schedule on receipt of the sealed envelopes at the Coordinating Center. The assignments in question were not reissued. The small amount of imbalance introduced in this way was not considered serious enough to justify the effort involved in reissuing the assignments.

The allocation schedule used by personnel in the CDP Coordinating Center revealed the contents of the bottles assigned (see Table 10-9). The presence of this information violates one of the masking safeguards listed in Table 10-8. However, there is no evidence that this information had any effect on the assignment process.

The mail system described was made possible

Table 10-9 Sample CDP treatment allocation schedule

Order of assignment within block	*Bottle number to be assigned*	*Bottle contents*	*Patient ID number*	*Patient name or name code*
1	29	CPIB	(56-001)	(JAMEI)
2	14	NICA	(56-002)	(ASJON)
3	26	PLBO	( )	( )
4	2	ESG2	( )	( )
5	27	ESG1	( )	( )
6	19	NICA	( )	( )
7	15	DT4	( )	( )
8	16	CPIB	( )	( )
9	13	PLBO	( )	( )
10	25	PLBO	( )	( )
11	10	ESG1	( )	( )
12	4	ESG2	( )	( )
13	24	PLBO	( )	( )
14	23	PLBO	( )	( )
15	9	DT4	( )	( )
16	30	ESG2	( )	( )
17	17	DT4	( )	( )
18	20	DT4	( )	( )
19	11	PLBO	( )	( )
20	6	CPIB	( )	( )
21	5	PLBO	( )	( )
22	28	ESG1	( )	( )
23	22	CPIB	( )	( )
24	18	ESG1	( )	( )
25	7	ESG2	( )	( )
26	8	NICA	( )	( )
27	1	PLBO	( )	( )
28	12	PLBO	( )	( )
29	21	PLBO	( )	( )
30	3	NICA	( )	( )

Source: Reference citation 104. Adapted with permission of the American Heart Association, Inc., Dallas, Texas.

Table 10-10 Sample CDP allocation form and envelope

Part A. CDP treatment allocation form

We have received your request for a treatment allocation for

Mr. ______________________________

whose identifying number is ______________________________

This person should receive medication from bottles identified by the following number:

[ ]

The sealed tear off portion of the label on each bottle should be removed prior to dispensing. The patient's name, treating physician, date and prescription number should be recorded on the tear off portion of the label prior to filing with the patient's prescription record.

The treatment should be initiated at Initial Visit 3 and should be administered on the following schedule:

1 capsule three times a day after meals from Initial visit 3 through Initial visit 4.

2 capsules three times a day after meals from Initial visit 4 through Initial visit 5.

3 capsules three times a day after meals after Initial visit 5 throughout the remainder of the study on the above named person unless clinically contraindicated

NOTE: If the date on which the treatment allocation envelope has been opened is more than four months after the date of Initial Visit 1 (which, as indicated on Form 01, is ______________), this allocation must be returned unused to the CDP Coordinating Center and this patient must start anew with Initial Visit 1.

Date of Allocation ______________________________

CDP Coordinating Center
Baltimore, Maryland 21201

Part B: Treatment allocation envelope

CORONARY DRUG PROJECT

Treatment Allocation for

Mr. ____________________ I.D. No. __________

DO NOT OPEN until instructed to do so in Form 02 (at Initial Visit 3).

If not opened within four months following the date of Initial Visit 1, this envelope should be returned to the CDP Coordinating Center.

Source: Coronary Drug Project Research Group.

because of the time separation between initiation of the request and treatment—generally about a month. Telephone assignments were allowed only when there was not adequate time to complete the mail circuit and then only if Coordinating Center personnel were satisfied that the patient in question was eligible for enrollment, that the clinic had completed the necessary forms, and that they had obtained his consent for enrollment.

The scheme described above cannot be used in cases where clinic personnel have to have the assignment as soon as the patient agrees to enroll. A system for making telephone allocations, such as used in CASS, has to be used in such cases, unless the study is willing to rely on a noncentral self-administered scheme (not recommended). The procedure in CASS required Coordinating Center personnel to carry out a series of telephone-administered checks with the requesting party before an assignment could be released. They included:

- Checks for eligibility
- Checks on the disease classification (needed for proper stratification)
- Checks to determine if the patient had signed the study consent statement and had indicated his willingness to accept either surgical or medical treatment.
- Checks to make certain a date for surgery had been set (for use if the patient was assigned to surgery)

CASS Coordinating Center personnel responsible for issuing assignments were masked with regard to assignments until the telephone interview was completed. This was done to protect against premature disclosure of assignments during the interview process.

The telephone assignment process used in CASS could be managed during the normal working hours of the Coordinating Center. This may not be possible in studies involving clinics scattered across a large number of time zones. Extended hours of phone coverage will be needed in such cases. Twenty-four-hour phone coverage will be needed when the trial involves emergency treatments that must be initiated as soon as possible.

The advent of low-cost, stand-alone minicomputers makes it possible to control the assignment process without any contact with the coordinating center, as in the Hypertension Prevention Trial (HPT). A clinic in that study initiated a request for assignment via an on-site computer (IBM S/23 DataMaster). The assignment was released via the computer, but only if the data forms entered by the clinic met the edit tests necessary for assignment.

Many trials, especially single-center trials, which cannot arrange for a centrally administered allocation scheme, must rely on self-administered schemes managed at the clinic. The usual approach in such cases is to place the assignments in sealed envelopes arranged in a predetermined order with personnel instructed to use the envelopes in order of arrangement, as indicated by numbers appearing on the faces of the envelopes. Strict ground rules should be established to indicate when envelopes are to be opened and to ensure that patients are counted in the trial once this has happened. Persons authorized to draw an allocation envelope should be required to check the prerandomization data form for missing data and for exclusion conditions before the envelope is opened. Documents completed in the allocation process should identify the patient for whom the assignment was intended and the time the envelope was opened. The time information is important when checks are made to determine if envelopes are used in the order indicated.

There is, of course, no method of allocation that is completely foolproof. It is important for this reason to perform periodic checks for breakdowns in the assignment process, regardless of how it is administered. It is dangerous to assume that the rules for allocation, no matter how explicitly outlined, will always be followed. The checking that is carried out should be performed by an individual or group of individuals not directly involved in the assignment process. For example, such checks in CASS were made by an external review team during visits to the CASS Coordinating Center. A similar function can be performed by the statistician or some other individual in the case of small-scale single-center trials using self-administered allocation schemes.

10.8 ILLUSTRATIONS

The illustrations in this section are designed to acquaint the reader with various techniques for constructing allocation schedules. The first 5 illustrations are for unmasked trials. Illustrations 6 and 7 are for masked trials. Illustration 1 involves use of random permutations of a set of numbers for constructing the randomization schedule. All of the remaining illustrations, except Illustration 5, involve use of random number tables. Illustration 5 involves use of a pseudo-random number generator.

10.8.1 Illustration 1: Restricted randomization using a table of random permutations

a. Specifications

- Treatment groups: 3
- Allocation ratio: 1:1:2
- Blocking constraints:
    - Number of blocks: 3
    - Block sizes: $k_1 = 12$, $k_2 = 4$, $k_3 = 4$
- Treatment masking: None
- Stratification variables: None
- Random permutation source: Cochran and Cox (1957). See Table 10-11.

b. Approach

Step 1 Establish treatment notation. Let:
T1 denote test treatment 1
T2 denote test treatment 2
C denote control treatment

Step 2 Establish treatment coding rule. Assign:
C for integers 1 through $k/2$
T1 for integers $1 + (k/2)$ through $3k/4$
T2 for integers $1 + (3k/4)$ through k

Step 3 Select a random start in table of random permutations. Set 7, Table 10-11, in this example.

Step 4 Establish reading rules. Read from left to right, i.e., use set 7 for first block, set 8 for second block, and set 9 for third block. Skip numbers in a permutation set that exceed the indicated block size.

Step 5 Record the assignment sequence. See third column of Table 10-12.

c. Comment

Note that the allocation ratio of 1:1:2 is satisfied in each of the three blocks.

10.8.2. Illustration 2: Unblocked allocations using a table of random numbers

a. Specifications

- Treatment groups: 2
- Allocation ratio: 1:1
- Blocking constraints: None

Table 10-11 Reproduction of 20 sets of random permutations of first 16 integers, from page 584 of Cochran and Cox (1957)

Permutation set																			
1	*2*	*3*	*4*	*5*	*6*	*7*	*8*	*9*	*10*	*11*	*12*	*13*	*14*	*15*	*16*	*17*	*18*	*19*	*20*
9	16	15	12	2	11	4	16	11	10	2	5	5	14	11	2	14	13	16	6
11	3	2	6	15	13	10	1	4	13	11	8	16	16	4	3	5	15	5	15
14	14	8	16	11	15	5	14	14	11	1	14	15	15	13	5	7	11	11	16
4	13	1	3	5	7	6	2	16	1	14	9	14	3	3	1	6	16	6	10
6	6	10	7	13	10	16	7	2	12	6	12	6	13	8	9	15	9	1	11
2	10	14	9	12	3	3	10	5	6	5	16	12	10	15	10	11	4	9	8
5	15	11	14	10	4	14	13	6	4	12	4	11	5	10	14	16	5	7	9
16	5	13	10	3	9	12	6	3	7	3	7	3	11	14	7	3	14	4	12
8	12	7	11	7	8	13	15	13	9	4	3	8	1	12	6	9	8	15	14
1	8	3	2	1	5	15	9	9	3	10	11	13	8	5	13	12	3	3	5
13	9	9	1	6	2	11	3	8	8	15	1	7	9	7	8	8	6	2	3
15	1	5	5	9	6	9	4	10	5	8	13	10	7	9	15	2	10	8	4
7	4	12	13	16	1	2	11	12	2	16	15	2	4	2	11	1	7	13	1
10	2	4	15	4	16	1	12	7	15	9	10	9	12	16	4	13	2	10	13
3	7	6	8	8	14	7	5	1	14	13	2	4	2	1	16	4	1	12	7
12	11	16	4	14	12	8	8	15	16	7	6	1	6	6	12	10	12	14	2

Source: Reprinted with permission of John Wiley & Sons, Inc., New York (copyright © 1957).

Table 10-12 Allocations for Illustration 1

Order of assignment		*Value from Table 10-11**	*Treatment assignment*
1	Block 1	4	C
2		10	T2
3		5	C
4		6	C
5		3	C
6		12	T2
7		11	T2
8		9	T1
9		2	C
10		1	C
11		7	T1
12		8	T1
13	Block 2	1	C
14		2	C
15		3	T1
16		4	T2
17	Block 3	4	T2
18		2	C
19		3	T1
20		1	C

*Starting point: Permutation set number 7, Table 10-11.

- Treatment masking: None
- Stratification variables: None
- Random number source: Rand Corporation (1955)

b. Approach

Step 1 Establish treatment codes. Let:
C denote control treatment
T denote test treatment

Step 2 Select an arbitrary starting point from table of random numbers. Suggested method:

i. Arbitrarily open book to some page and place the point of a pencil on the page without looking. Use the three digits to immediate right and nearest the point to designate the starting page (17 in example).
ii. Repeat the process described in step i to select a starting column (22 in example) and row (3 in example) for the page selected (page 17 in the example, see Table 10-6).

Step 3 Define order in which numbers are to be used. Read from left to right and down by row, beginning at the point designated in Step 2. Use single integers.

Step 4 Establish correspondence between numbers selected and treatment assignments. For this illustration use odd integers (1, 3, 5, 7, 9) to designate assignment to the control treatment (C) and even integers (0, 2, 4, 6, 8) to designate assignment to the test treatment (T).

Step 5 Record the treatment assignment sequence (see third column of Table 10-13).

c. Comment

Note that the sequence for the first 20 assignments provided 9 T assignments and 11 C assignments for an observed allocation ratio of 1:1.2 instead of the desired ratio of 1:1.

10.8.3 Illustration 3: Blocked allocations using the Moses-Oakford algorithm and a table of random numbers

a. Specifications

Same as for Illustration 2 except:

- Blocking constraints:
  - Number of blocks: 4
  - Block sizes: $k_1 = 10$, $k_2 = 4$, $k_3 = 2$, $k_4 = 4$

Table 10-13 Allocations for Illustration 2

Order of assignment	*Random number**	*Treatment assignment*
1	8	T
2	7	C
3	9	C
4	4	T
5	9	C
6	6	T
7	7	C
8	9	C
9	9	C
10	6	T
11	7	C
12	5	C
13	8	T
14	8	T
15	7	C
16	5	C
17	0	T
18	4	T
19	9	C
20	8	T

*Starting point. Row 3, Column 22, Table 10-6.

b. Approach

Step 1 Same as for Illustration 2.

Step 2 Starting point: row 9, column 42, Table 10-6.

Step 3 Reading instructions: Left to right to end of row, then down, row by row. Use pairs of integers as long as the remaining block size is ≥10. Skip 00 and pairs of integers that exceed remaining block size. Use single integers once the remaining block size is ≤9. Ignore 0. (*Note:* Most numbers exceeding the remaining block size could be converted to the usable range through subtraction of an appropriate multiplier of the remaining block size if desired. For example, the number 53 converts to 9 by subtracting 44 if the remaining block size is 11. However, such arithmetic is tedious and subject to error if done by hand and therefore is not done in this example.)

Step 4 Set down an arbitrary order of treatments, as shown in column 2, Table 10-14.

Step 5 Establish the final order of treatment assignment (column 4, Table 10-14) using the Moses-Oakford algorithm (Table 10-3).

c. Comment

The table below gives the location of the first and last numbers used from Table 10-6 for each of the four blocks.

Block number	*First number*		*Last number*	
	Column	*Row*	*Column*	*Row*
1	42	9	22	10
2	31	10	40	10
3	50	10	50	10
4	3	11	14	11

10.8.4 Illustration 4: Stratified and blocked allocations using the Moses-Oakford algorithm and a table of random numbers

a. Specifications

- Treatment groups: 3
- Allocation ratio: 1:1:5
- Blocking constraints: Blocks of sizes 7 or 14 arranged in random sequence

Table 10-14 Allocations for Illustration 3

No. of blocks	Treatment codes	Random numbers*	Page	Column	Row
4	C = Control	Start	17	42	9
Block sizes: 10, 4, 2, 4	T = Test	End	17	14	11
Allocation ratio: 1:1		Source: Rand Corp. (1955)			

(1a) Order of assignment		(1) Order of assignment within block	(2) Treatment assignments: Initial	(3) Replacements	(4) Final	(5) Random number
1	Blk 1	1	C	—	C	—
2	Blk 1	2	~~T~~	T	T	2
3	Blk 1	3	~~C~~	~~T~~, ~~C~~, ~~C~~, T	T	2
4	Blk 1	4	~~T~~	~~C~~, T	C	3
5	Blk 1	5	C	—	C	3
6	Blk 1	6	T	—	C	4
7	Blk 1	7	C	—	T	3
8	Blk 1	8	~~T~~	T	C	3
9	Blk 1	9	C	—	T	4
10	Blk 1	10	T	—	T	08
11	Blk 2	1	~~C~~	~~T~~, T	T	—
12	Blk 2	2	T	—	T	1
13	Blk 2	3	C	—	C	3
14	Blk 2	4	T	—	C	1
15	Blk 3	1	~~C~~	T	T	—
16	Blk 3	2	T	—	C	1
17	Blk 4	1	C	—	C	—
18	Blk 4	2	~~T~~	C	C	2
19	Blk 4	3	C	—	T	2
20	Blk 4	4	T	—	T	4

* Reading rule:
See Illustration 3 (Section 10.8.4) for reading instructions. Note also that the last two assignments in block 4 could be made without drawing any number since the two remaining assignments must be to the control treatment in order to satisfy the allocation ratio.

- Treatment masking: None
- Stratification variables: 1 (2 levels)
- Random number source: Rand Corporation (1955)

Step 1 Establish treatment codes (shown in Table 10-15)

Step 2 Starting point: row 10, column 6, Table 10-6.

Step 3 Reading rule: Read left to right using pairs of integers until remaining number of assignments to be made is ≤9, then use single integers. Ignore 00 and 0.

Step 4 Set down an arbitrary order of treatments, as shown in column 2, Table 10-15.

Step 5 Establish the final order of treatments (column 4, Table 10-15) using the Moses-Oakford algorithm (Table 10-3).

c. Comment

This problem requires construction of an allocation schedule for enrollment of an unspecified number of patients per stratum (See Chapter 14 for discussion of difficulties with designs involving recruitment quotas). One approach would be to develop 2 schedules, each one with enough allocations to meet the recruitment goal of the trial. However, this is wasteful since half of the assignments would remain unused. A more effi-

Table 10-15 Allocations for Illustration 4

Block number	Treatment Codes	Random numbers*	Page	Column	Row
3	C = Control	Start	17	11	10
Block size: 14	T1 = Test trt 1	End	17	33	11
Allocation ratio: 1:1:5	T2 = Test trt 2	Source: Rand Corp. (1955)			

Patient ID number	(1a) Order of assignment	(1) Order of assignment within stratum	(2) Treatment assignments: Initial	(3) Replacements	(4) Final	(5) Random number
—	—	1	~~C~~	~~T1~~, ~~C~~, ~~C~~, C	C	—
—	—	2	~~C~~	~~T2~~, ~~T1~~, ~~C~~, ~~C~~, C	C	1
—	—	3	~~C~~	~~T2~~, C	C	1
—	—	4	~~C~~	C	C	2
—	—	5	C	—	C	4
—	—	6	C	—	T1	1
—	—	7	~~C~~	C	T2	3
—	—	8	C	—	C	2
—	—	9	C	—	T1	2
—	—	10	C	—	C	07
—	—	11	T1	—	T2	02
—	—	12	T1	—	C	01
—	—	13	T2	—	C	03
—	—	14	T2	—	C	02

* Reading rule:
See Illustration 4 for reading instructions (Section 10.8.4).

cient approach is to generate sets of worksheets arranged in the order generated as dictated by the random sequence of block sizes used (in this example, 7, 7, 14, 7, 14, etc.; see Table 10-15 for third block of size 14). They are then used in order, as needed, depending on enrollment patterns in the 2 strata. The first worksheet of block size 7 is used to make assignments for patients in the stratum represented by the first patient enrolled into the trial. For example, if the stratification variable is sex and the first patient enrolled is female, then the first worksheet is used for the first 7 females enrolled. The second sheet is not used until the eighth patient enters the same stratum, or until a patient enters who qualifies for the second stratum, in this example, a male. The stratum number is not placed on the sheet until it is used. The lines in the column labeled *Patient ID number* would be filled in as the individual assignments are issued. The numbers written in column 1a would depend on the number of sheets already used for allocations to the stratum in question. For example, they would run from 1 through 14 if there had been no previous assignments in the stratum, and from 8 through 21 if a block of size 7 had already been filled for the stratum.

10.8.5 Illustration 5: Sample allocation schedule for the Macular Photocoagulation Study using pseudo-random numbers

a. Specifications

- Treatment groups: 2
- Allocation ratio: 1 : 1
- Blocking constraints: Blocks of sizes 6 or 8 in random sequence
- Treatment masking: None
- Stratification variables: 2 (clinic and type of eye disease, three different types)
- Random number source: Computer based pseudo-random number generator

b. Approach

Step 1 Establish treatment codes. Let:
C denote control treatment
T denote test treatment

Step 2 Select a block size, 6 or 8, by some random or pseudo-random process (size 6 in this Illustration).

Step 3 Arrange treatment codes in arbitrary order (column 2, Table 10-16).

Step 4 Generate a sequence of 5-digit pseudo-random numbers and record in the order generated (column 3, Table 10-16).

Step 5 Link the treatment code (column 2, Table 10-16) and pseudo-random number (column 3, Table 10-16).

Step 6 Order the pseudo-random numbers with associated treatment codes (column 4, Table 10-16).

Step 7 Repeat steps 2 through 6 as necessary to generate the desired number of assignments.

c. Comment

Sheets should be used in the order needed, as discussed in Illustration 4.

10.8.6 Illustration 6: Double-masked allocation schedule using the Moses-Oakford algorithm and a table of random numbers

a. Specifications

- Treatment groups: 3
- Allocation ratio: 2 : 2 : 3
- Blocking constraints: Uniform block size of 14
- Treatment masking: Double-masked
- Stratification variables: 1 (2 levels)
- Random number source: Rand Corporation (1955)

b. Approach

The first step is to denote the bottle numbers to be used. The designation in this Illustration was made by arbitrarily selecting a random permutation of the first 16 integers (set 12, Table 10-11). The first 6 values of size 14 or less in the permutation are used to denote bottles containing the control drug, the next 4 numbers are used to designate bottles containing test drug 1 and the last 4 numbers are used to denote bottles containing test drug 2. The bottle codes and associated treatment are recorded in column 2, Table 10-17, and then rearranged as described for Illustrations 3 and 4 to yield the bottle sequence indicated in column 6. The sheet provided is for 1 block in the scheme.

c. Comment

Note that each bottle number appears only once in Table 10-17. Subsequent blocks will contain different orderings of the same bottle numbers.

Table 10-16 Sample allocation schedule from the Macular Photocoagulation Study for Illustration 5

(1) *Order of assignment in block*	(2) *Initial assignment*	(3) *Pseudo-random number*	(4) *Ordered pseudo-random number with treatment code*	(5) *Final assignment from column 4*
1	T	26391	(T) 07631	T
2	C	29126	(C) 10645	C
3	T	07631	(C) 22846	C
4	C	22846	(T) 26391	T
5	T	30856	(C) 29126	C
6	C	10645	(T) 30856	T

Table 10-17 Allocation schedule for double-masked drug trial described in Illustration 6

Block size: 14

Allocation ratio: 2:2:3

Treatment codes:
C: Bottle Nos. 4, 5, 8, 9, 12, 14
T1: Bottle Nos. 1, 3, 7, 11
T2: Bottle Nos. 2, 6, 10, 13

Random numbers*

	Page	Column	Row
Start	17	21	7
End	17	19	8

Source ______

(1) Order of assignment within block	(2) Treatment assignments: Initial	(3) Treatment assignments: Replacements	(4) Treatment assignments: Final	(5) Random number	(6) Bottle number
1	~~C-4~~	~~T1-1~~, T2-6	T2-6	—	6
2	~~C-5~~	T2-10	T2-10	2	10
3	~~C-8~~	T2-10	C-5	2	5
4	~~C-9~~	T1-11	T1-11	4	11
5	~~C-12~~	T2-6	T1-1	1	1
6	C-14		C-14	6	14
7	T1-1		C-4	1	4
8	T1-3	~~T2-13~~, T2-6	C-12	5	12
9	T1-7	T1-11	C-9	4	9
10	T1-11		T1-7	09	7
11	~~T2-2~~	T2-10	C-8	03	8
12	T2-6		T2-13	08	13
13	T2-10		T2-2	11	2
14	T2-13		T1-3	08	3

* Reading rule:
Read numbers from left to right and down, row by row. Use pairs of numbers as long as the remaining block size is >10. Skip 00's and pairs of numbers which exceed the remaining block size. Use single integer once the remaining block size is <9. Skip 0's

10.8.7 Illustration 7: Sample CDP double-masked allocation schedule

a. Specifications

- Number of assignments: 8,341
- Clinics: 53
- Treatment groups: 6
- Allocation ratio: 2:2:2:2:2:5
- Blocking constraints: Uniform block size of 15
- Treatment masking: Double-masked
- Stratification variables: 2 (clinic and risk group, two levels per clinic, to yield a total of $53 \times 2 = 106$ allocation strata)
- Random number source: Rand Corporation (1955)

b. Approach

The allocation procedure is described in a CDP publication (Coronary Drug Project Research Group, 1973a). Treatment assignments were identified by a 2-digit bottle number as shown in Table 10-9. The same bottle numbers were used in all clinics. Hence, all bottles bearing a particular number always contained the same medication, regardless of clinic.

c. Comment

Note that each bottle number appears once and only once in the 30 assignments listed in Table 10-9 and that both blocks in the table satisfy the allocation ratio (i.e., contain 2 assignments to each test treatment and 5 assignments to the placebo treatment).

11. The study plan

The way to improve a treatment is to eliminate controls.

Hugo Menuch

11.1 INTRODUCTION

The basic elements of the plan for any trial will be set long before the first patient is enrolled. The nature of the test treatment and outcome measure will be specified in the funding proposal. Specifics having to do with execution of the study plan may not be addressed until the trial has been funded. The period of time between initiation of funding and enrollment of the first patient is one requiring intense effort to develop and test procedures needed for the trial. However, the planning and testing process does not end there. In fact, it is likely to continue over much of the course of the trial, particularly in long-term trials involving extended periods of patient recruitment or follow-up. The goal in such settings of maintaining the study plan unchanged once the first patient has been enrolled, while laudable, is not always practical.

The term *study plan* used in a broad sense refers to the design of the trial and all the organizational and operational details needed to carry it out. In this sense, various other chapters, in addition to this one, relate to the study plan, starting with the two previous chapters and including most of those that follow.

11.2 DESIGN FACTORS AND DETAILS TO BE ADDRESSED IN THE STUDY PLAN

No trial should be undertaken without:

- A concise statement of its objective(s)
- A specification of the outcome measure(s) to be used for evaluating the study treatments
- Agreement on the treatments to be tested
- A sample size calculation that indicates the required number of patients, or a calculation of the power provided with a prestated sample size
- Specification of the required length of patient follow-up
- A specified set of patient entry and exclusion criteria
- A method for randomization
- A specified baseline and follow-up examination schedule
- A set of data intake procedures, including specification of the methods for data entry, editing, and quality control
- An established organizational and decision-making structure

Agreement on the design and operating features of a trial cannot be ensured unless they have been written down and have been reviewed and accepted by investigators responsible for the trial.

11.3 OBJECTIVE AND SPECIFIC AIMS

The statement of the primary objective is by far the most important specification in the trial. It must be formulated and agreed upon before a data collection scheme can be developed. The statement should indicate the:

- Type of patients to be studied
- Class of treatments to be evaluated
- Primary outcome measure

Sample statements of objectives follow:

University Group Diabetes Program (UGDP)

Evaluation of the efficacy of hypoglycemic treatments in the prevention of vascular complications in a long-term, prospective, and cooperative clinical trial (University Group Diabetes Program Research Group, 1970d).

Coronary Drug Project (CDP)

Evaluate the efficacy of several lipid-influencing drugs in the long-term therapy of CHD in men ages 30 through 64 with evidence of previous myocardial infarction (Coronary Drug Project Research Group, 1973a).

National Cooperative Gallstone Study (NCGS)

To determine the efficacy of oral administration of a high and low dose of CDC acid in dissolving or reducing the size of cholesterol gallstones, as compared with placebo treatment (National Cooperative Gallstone Study Group, 1981a).

The statement from the CDP comes closest to satisfying the three requirements stated above. It indicates the type of patients to be treated and the class of treatments to be used. However, it is ambiguous with regard to outcome, other than to suggest that it is related to coronary heart disease (CHD). The UGDP statement indicates nothing about the study population and is ambiguous with regard to chosen outcome measure. The NCGS statement names the treatment and outcome measure, but says nothing about the study population.

It is not uncommon for a large-scale trial to have secondary objectives as well. They are illustrated for the three trials cited above.

UGDP

- To study the natural history of vascular disease in maturity onset, noninsulin dependent diabetics.
- To develop methods applicable to multicenter clinical trials.

CDP

- To obtain information on the natural history and clinical course of CHD.
- To develop more advanced technology for the design and conduct of large, long-term, collaborative clinical trials.

NCGS

- To determine whether either a high or low dose of chenodeoxycholic acid could be safely used to dissolve cholesterol gallstones.
- To determine the rate of recurrence of gallstones in those patients in which chenodeoxycholic acid feeding has successfully dissolved gallstones.

Whenever multiple objectives are stated, it is wise to rank them in order of importance. The ranking will have important design implications especially if data requirements for the objectives differ. The investigators should state the specific aims to be pursued in conjunction with each objective. The methods and data collection requirements of the trial should then be constructed to satisfy the stated aims.

11.4 THE TREATMENT PLAN

General considerations involved in choosing the test and control treatments were discussed in Chapter 8. Once they have been selected, it is necessary for investigators to address a series of practical issues concerning treatment administration. One issue in drug trials concerns whether the treatments are to be administered using a fixed- or variable-dosage schedule. Ideally, the administration schedule should be as near that used in actual practice as feasible. A variable dosage schedule, tailored to the needs of individual patients, should be used if the test drug is ordinarily used in this way. A fixed-dosage schedule may be used if the drug is normally used in this way or if the variation in dosages used is small.

The choice may be constrained by masking requirements. The desire to individualize treatment in order to achieve some desired effect (for example, to normalize blood glucose levels in the case of a hypoglycemic drug) may have to be

abandoned if there is to be double-masked administration of the treatments. The manipulations required for dosage titrations can be hazardous to patients if they are done in a masked fashion and may in any case render the masking ineffective.

Another issue in drug trials has to do with the formulation of the test treatment. Whenever feasible, it should be used in the same form as in normal practice. However, here again some compromises may be necessary. For example, investigators may choose to use capsules for dispensing study medications even though the test drug is normally dispensed in tablet form in order to mask the taste and appearance of the study drugs. Modification in the form or route of administration is acceptable only if it does not affect the bioavailability or pharmacological action of the study drugs.

A key design decision in trials involving two or more treatments that may be used alone or in combination concerns whether a factorial treatment structure should be used (see Glossary for definition). Table 11-1 illustrates use of this design for a two-drug study. Separate placebos for each drug tested are necessary when the test drugs are to be dispensed on different time schedules or in different forms (e.g., capsules for one test drug and tablets for the other test drug). Patients in the cell designated AB would receive both drug A and B, those in cell $A\bar{B}$ would receive drug A and the placebo for drug B, and so on. A trial involving three different drugs, each administered at a single, fixed-dose level and suitable for use alone or in combination, would involve eight (i.e., 2^3) treatment combinations: ABC, $AB\bar{C}$, $A\bar{B}C$, $\bar{A}BC$, $A\bar{B}\bar{C}$, $\bar{A}B\bar{C}$, $\bar{A}\bar{B}C$, and $\bar{A}\bar{B}\bar{C}$.

The main advantage of a factorial treatment structure lies in the opportunity it provides for estimating both individual and combined treatment effects via the same experiment. A full factorial treatment structure (see Glossary for definition) should be considered whenever there is a reason to suspect additive or synergistic treatment effects. It should not be used with treatments that are incompatible, or where there is no interest in some of the treatment combinations. A partial factorial treatment structure (see Glossary) may be considered in the latter case, as in the Persantine Aspirin Reinfarction Study (PARIS). Persantine was not used alone, because of the high dose level required in the absence of aspirin and because of previous animal work suggesting that the combination of aspirin and persantine had a more profound effect on blood platelets than either drug alone (Persantine Aspirin Reinfarction Study Group, 1980b). The primary aim of the study was to provide a comparison of the combination of persantine and aspirin against aspirin alone. A secondary aim was to measure the usefulness of this combination against a placebo treatment. This difference in interest is reflected by the fact that the number of patients assigned to the placebo treatment was only half the number assigned to either of the other two treatment groups (Table 11-2).

The methods for administering the treatments and ground rules under which treatments may be altered or stopped should be set down in the treatment protocol. Table 11-3 provides a list of the items that should be included in this document.

The details of the protocol should be subjected to careful review before implementation. Lack of agreement can lead to unacceptable variation in the data collection or treatment process. Establishing standards for data collection and treatment administration is important whenever multiple investigators are involved in a trial, whether they are located in a single clinic or in multiple clinics.

Medical conditions that may require a study physician to depart from the assigned treatment should be detailed. It is also wise to outline side effects that are a normal part of a drug's pharmacological effect. For example, the treatment protocol for the NCGS warned physicians

Table 11-1 Example of a factorial treatment design for a two-drug study

Drug	B	$\bar{B}$
A	AB	$A\bar{B}$
$\bar{A}$	$\bar{A}B$	$\bar{A}\bar{B}$

Table 11-2 Numbers of patients by treatment group in PARIS

Drug	*Persantine*	*Persantine placebo*
Aspirin	810	810
Aspirin placebo	0	406

Table 11-3 Major items to be included in the treatment protocol

- Specification of the test and control treatments to be tested and rationale for the choices
- Review of previous research on the safety and efficacy of the proposed treatments
- Description of the methods for administering the test and control treatments
- List of contraindications for the proposed treatments
- Specification of the clinical conditions that may necessitate termination of the assigned treatment
- Specification of side effects that may require termination of the assigned treatment, as well as those that should not
- Methods, in the case of masked drug trials, for packaging and dispensing drugs, including a general outline of the conditions under which the masking may have to be revealed to clinic personnel or to a study patient
- General scheme to be used for assigning patients to the study treatments

against stopping a patient's treatment because of mild diarrhea, since such problems were a recognized side effect of chenodeoxycholic acid therapy and were not considered to be serious (National Cooperative Gallstone Study Group, 1981b).

The conditions under which a treatment assignment is revealed to clinic personnel in a double-masked trial should be specified. As a rule, there are few valid reasons for unmasking assignments during the course of the trial, since the assigned treatments can be terminated without revealing their identity to patients or clinic personnel. For example, provisions for unmasking in the CDP were limited to emergencies involving life-threatening uses of a medication by a patient or a member of his family or when a patient required emergency surgery and the surgical team needed to know his treatment assignment. Patients undergoing elective surgery simply stopped taking their study medicine before the surgery and during the recovery period.

11.5 COMPOSITION OF THE STUDY POPULATION

The formulation of patient selection criteria for the study represents a balance of two opposing forces: one designed to produce a highly homogeneous study population and the other designed to minimize the restrictions on the study population and hence maximize the opportunities for patient recruitment. On the one hand the more homogeneous the population, the more precise the study, and hence the smaller the number of patients needed to detect a given difference. On the other hand, the greater the heterogeneity, the broader the basis for generalizing findings at the end of the study. The advantages and disadvantages of different selection strategies are summarized in Table 11-4.

Investigators must agree on selection and exclusion criteria before patient recruitment starts. Often they fail to appreciate the impact the criteria will have on recruitment. Estimates of patient availability made during the design stage of the trial are likely to be unrealistically high unless they are based on actual patient surveys using the proposed criteria. Factors that are not likely to influence outcome should not be used for exclusion, since they do nothing to improve the precision of the trial while they make patient recruitment more difficult. Table 11-5 lists the main selection criteria used in the trials sketched in Appendix B.

Socioeconomic status is usually not a valid basis for patient selection. Neither the scientific nor the lay community is likely to look kindly on such forms of selection. Selection on the basis of ethnic origin, religion, or race should also be

Table 11-4 Advantages and disadvantages of opposing selection strategies

Highly restrictive selection criteria
- **Advantages**
  - Provides more precise comparison of the test and control treatments
  - Results of the trial less likely to be effected by population variability
- **Disadvantages**
  - Increases the cost and time required for patient recruitment
  - Limits the generalizability of the study findings

Minimally restrictive selection criteria
- **Advantages**
  - Makes patient recruitment easier
  - Provides base for wider generalization of findings
- **Disadvantages**
  - May obscure treatment effects because of variability in composition of study population
  - Results of the trial may be confusing, especially if an observed effect appears to be associated with a subgroup of patients in the study and the subgroup is too small to yield a reliable treatment comparison

Table 11-5 Primary selection criteria of trials sketched in Appendix B

Trial	*Sex*	*Age limits on entry*	*Disease state*
AMIS	Both	30-69	Prior MI
CASS	Both	None	Prior MI
CDP	Males	30-64	Prior MI
HDFP	Both	30-69	Diastolic blood pressure ≥95 mm Hg
HPT	Both	25-49	Diastolic blood pressure ≥78 but <90 mm Hg
IRSC	Both	<10	Grade III or IV vesicoureteral reflux
MPS	Both	≥50 for SMD, ≥18 for HISTO, None for INVM	Evidence of neovascularization for all three conditions
MRFIT	Males	35-57	High risk for CHD
NCGS	Both	None	Radiolucent gallstones
PARIS	Both	30-74	Prior MI
PHS	Males	40-75	Absence of MI history
POSCH	Both	30-64	Hypercholesterolemia
UGDP	Both	None	Newly diagnosed diabetes
VACSP 43	Males	None	Evidence of gangrene of either foot

avoided. A possible exception relates to diseases or conditions concentrated primarily, if not exclusively, in individuals of a particular religious, ethnic, or racial background. However, even if one avoids use of such factors, the study population of a clinic may be quite homogeneous with regard to them. The socioeconomic, ethnic, or racial spectrum covered by a study population will be a function of where and how it is recruited. The racial mix of clinics in the UGDP varied from being nearly all white to being nearly all black (University Group Diabetes Program Research Group, 1970e). This variation stands in marked contrast to that observed in the Coronary Artery Surgery Study. The population in that study was virtually all white—a reflection, undoubtedly, of the nature of the patients served by the participating clinics and of the popularity of bypass surgery in white middle-class America (Coronary Artery Surgery Study Research Group, 1981).

Of the 14 trials listed in Table 11-5, 4 used sex as an exclusion. The sex restriction in the CDP was required because estrogen—one of the drugs tested in that trial—was contraindicated for females. The Veterans Administration Cooperative Study Program No. 43 (VACSP 43) and the Physicians' Health Study (PHS) excluded females from enrollment simply because of the small number of females contained in the populations approached for study. The rationale for the restriction in the Multiple Risk Factor Intervention Trial (MRFIT) is less clear. There is no question that even if the trial had been open to females that the majority of enrollees would have been male. However, that fact alone does not provide a sufficient rationale for the exclusion. Valid treatment comparisons can be made so long as the proportionate mix of males and females is the same across study treatments.

Ten of the 14 trials used age as a selection criterion. Generally, practical considerations figured in the limits used. For example, this was the case in the choice of the lower age limit for the Hypertension Prevention Trial (HPT). The original design called for a lower limit of 18. Ultimately, however, the limit was raised to 25 before the study started because problems were anticipated in recruiting and following people aged 18 to 25.

The use of upper age limits, especially in studies involving adult populations, is less easy to justify. CDP investigators arbitrarily imposed an upper limit of 65 primarily as a means of excluding individuals who had experienced their first MI relatively late in life. The limit made recruitment more difficult and in all probability did little to improve the precision of the trial, since there is no reason to believe the study treatments are any more or less effective in individuals over 65 than for those under 65.

11.6 THE PLAN FOR PATIENT ENROLLMENT AND FOLLOW-UP

The study plan should include a description of methods to be used for patient recruitment and an outline of the data collection schedule (see Chapters 12 and 14). Ideally, there should be at least two separate patient contacts before randomization with adequate time between the contacts to:

- Allow clinic staff time to consider the suitability of the patient for study
- Facilitate the identification of "faint of heart" patients
- Allow a staged approach to the informed consent process (see Section 14.6)

The study design should provide for a landmark that when passed marks entry of a patient into the trial (e.g., the point at which the treatment assignment is divulged to clinic personnel). A patient should be counted as part of the study population, regardless of his subsequent course of treatment, once the landmark has been passed.

After enrollment, patients will be required to return for one or more scheduled follow-up visits. The timing of these visits will depend on the data collection requirements of the study. The frequency is usually highest right after the initiation of treatment. The CDP required a clinic visit of each patient at one month and again at two months after enrollment for dosage increases. The next required visit was at four months after enrollment and then every four months thereafter (Coronary Drug Project Research Group, 1973a).

Except in special cases, the frequency of required data collection visits should be the same for all patients. A difference in the visit rates can bias the study results if it influences the rate at which clinical events are diagnosed and reported. This kind of bias was of concern in the Hypertension Detection and Follow-Up Program (HDFP) because of more frequent contacts with patients assigned to stepped-care than with those assigned to usual care (Hypertension Detection and Follow-Up Program Cooperative Group, 1979a).

The possibility of bias is not eliminated by use of identical schedules for required visits if the rate of interim unscheduled visits between scheduled visits is different for the study groups. A differential rate of unscheduled observations can still bias the way in which events are diagnosed and reported in the trial. Most long-term trials keep track of such contacts, if for no other reason than to provide a means of comparing the study groups for differences in contact rates.

The study plan should include provision for some minimal form of follow-up for dropouts (see Glossary for definitions). The follow-up may be for mortality only or for other kinds of outcomes, depending on the trial. See Chapter 15 for more details.

11.7 THE PLAN FOR CLOSE-OUT OF PATIENT FOLLOW-UP

An important design issue concerns disengagement of a patient from the trial when it is finished. Two general models are used for this purpose. One model is characterized by a common closing date for all patients, regardless of the date of enrollment. Another involves close-out after a specified length of follow-up. The latter approach requires as much time for close-out as for enrollment, whereas close-out takes place at the same time for all patients, regardless of when they were enrolled, when the former approach is used (see Section 15.4 for added discussion).

The CDP is an example of a trial using a common close-out date. All patients were separated from the study during June through August of 1974 (Coronary Drug Project Research Group, 1975). The NCGS provides an example of close-out after a specified period of follow-up—two years (National Cooperative Gallstone Study Group, 1981a).

12. Data collection considerations

Investigators seem to have settled for what is measurable instead of measuring what they would really like to know.

Edmund D. Pellegrino

12.1 INTRODUCTION

Decisions regarding the data collection schedule and related forms are among the most important in the trial. They will determine both the amount and quality of data generated in the trial.

There must be adequate time, once the study is funded and before data collection starts, for investigators to agree on the details of the data collection process. They must be concerned first with setting the schedule at which patients are seen, both before and after entry into the trial, and then with outlining the specific items of information to be collected each time the patient is seen. The investigators should allow adequate time after these steps are completed for developing and testing required data forms and for receiving and reacting to suggestions from clinic personnel who must use them.

The form development process should be undertaken by personnel who are experienced in form construction and who are familiar with methods for data collection and data processing in prospective studies. The development of data forms can be facilitated by review of sample forms used in other trials, especially those from trials with design and operating features similar to the one in question. Some of the desired samples can be obtained through the published literature (e.g., see appendixes in Coronary Drug Project Research Group, 1973a, and Coronary Artery Surgery Study Research Group, 1981) or via a central respository (e.g., see National Cooperative Gallstone Study Group, 1981a, for reference to forms placed on file at the National Technical Information Service). Others will have to be obtained by direct request to investigators involved in the trials of interest.

The reference list in Appendix I includes a number of citations pertinent to data collection and the construction of forms. Several of the references are from interview and survey literature but are relevant to clinical trials as well. A classic book by Payne (1951), although focused on opinion polling, is useful reading for anyone involved in data collection. The *Teacher's Word Book of 30,000 Words* (Thorndike and Lorge, 1944) indicates the expected level of comprehension of words as a function of education level. It is a useful resource, especially when forms are being designed for use in patient interviews.

Also included are several textbooks with chapters on forms design (Backstrom and Hursh-César, 1981; Kidder, 1981; Marks, 1982; Sudman and Bradburn, 1983), as well as a number of journal articles. The three articles by Wright and Haybittle (1979a,b,c) and a chapter from a monograph from the Coronary Drug Project (Knatterud et al., 1983) have direct relevance to the field of clinical trials. Papers by Collen and co-workers (1969), Helsing and Comstock (1976), Hochstim and Renne (1971), Holland and co-workers (1966), and Milne and Williamson (1971) deal with data collection via questionnaires. Other papers of interest include those by Barker (1980), Barnard et al. (1979), Bishop et al. (1982), Duncan (1979), Edvardsson (1980), Finney (1981), Layne and Thompson (1981), McFarland (1981), Romm and Hulka (1979), Roth et al. (1980), Schriesheim (1981), Smith (1981), and Zelnio (1980).

12.2 FACTORS INFLUENCING THE CLINIC VISIT SCHEDULE

12.2.1 Introduction

Every clinical trial must provide for data collection at a minimum of two time points: at or just before randomization and the initiation of treatment to provide baseline data, and at least once after randomization for collection of follow-up data. It is possible to collect all the required data for a patient during a single clinic visit if it is possible to collect the necessary baseline data, issue the treatment assignment, administer the treatment, and make the required follow-up observations all on the same day. However, the usual situation is one in which a patient is required to make one, two, or even more visits to the clinic on different days before he or she can be enrolled and assigned to treatment. Thereafter, the patient may need to make a series of return visits, extending over a period of weeks, months, or even years, to receive the assigned treatment and for follow-up data collection.

The discussion throughout this book deals with trials in which data collection is performed on an outpatient basis. If any hospitalization is required, it is assumed to be a small portion of the total time the patient is expected to be under study.

Patient visits that take place before the randomization visit are herein referred to as prerandomization visits. Enrollment into the trial occurs at the randomization visit and is marked by some explicit act (e.g., the opening of the treatment allocation envelope). Thereafter, the patient is a member of the treatment group to which he or she was assigned.

It is conventional to consider data collected at the prerandomization and randomization visits as baseline data and to refer to both types of visits as baseline visits (see Glossary). This convention will be followed in this book. It is reasonable if all data collected at the randomization visit are collected before initiation of treatment. Post-randomization visits include all visits that take place after the randomization visit. All such visits will be referred to as follow-up visits in this book, whether they are done on a scheduled or ad hoc basis.

12.2.2 Baseline clinic visit schedule

Baseline visits (prerandomization and randomization visits) are needed to:

- Determine a patient's eligibility for enrollment
- Provide baseline data for assessing changes occurring after the initiation of treatment
- Explain the purpose of the study to the patient and to obtain consent for participation in the trial
- Issue the treatment assignment

Whenever possible, it is useful to have the patient make at least two visits to the clinic before enrollment. The visits may be only a few days apart, especially when there is an urgent need to initiate treatment, or they may extend over a period of weeks, or even months. The repeat visits make it possible to replicate certain key baseline measurements. A time separation between visits may be needed as well to:

- Perform the necessary screening and diagnostic procedures for determining patient eligibility
- Allow sufficient time for a patient to recover from a procedure performed at one visit and to go through the preparatory steps required for the next clinic visit
- Provide adequate time for the informed consent process
- Allow adequate time for clinic staff to evaluate the data collected on the patient before enrollment

The Coronary Drug Project (CDP) required two prerandomization visits. The first visit was used to make an initial determination of a patient's eligibility for enrollment into the study, to obtain serum for lipid and other determinations, to perform a general physical examination, and to provide the patient with a preliminary explanation of the study. The second visit, scheduled approximately 1 month after the first visit, was used to assess a prospective patient's adherence to the prerandomization treatment schedule,[1] to obtain additional serum for a repeat set of laboratory determinations, and to obtain the patient's signed consent to participate in the trial. The randomization visit, scheduled approximately 1 month after the second prerandomization visit, was used for a final asessment of the patient's suitability for enrollment into the trial, including a further assessment of his adherence to the assigned medication schedule, and verification that the patient was indeed willing to be randomized. If so, the treatment allocation envelope was opened and the assigned treatment was initiated (Coronary Drug Project Research Group, 1973a).

Required diagnostic and data collection procedures should be designed to minimize patient inconvenience and exposure to unnecessary procedures, particularly those entailing risks to the patient. Hence, whenever feasible, the simplest procedures with the least risk should be performed first so that patients who then prove to be ineligible can be spared the inconvenience (and risks, if any) of the more complex and time-consuming procedures.

12.2.3 Follow-up clinic visit schedule

A follow-up visit is any visit, either required or nonrequired, to the study clinic by a patient who has been enrolled into the trial (i.e., assigned to treatment) that takes place after the randomization visit. Required visits should be specified in the study protocol and should be scheduled to take place at specified time points after the randomization visit. They are herein variously referred to as scheduled follow-up visits, required follow-up visits, or, in contexts where the meaning is clear, simply as follow-up visits. Visits of this class are needed to:

- Carry out procedures specified in the study protocol, including those for treatment administration and treatment adjustment
- Evaluate the patient's response to treatment
- Assess patient and physician adherence to the assigned treatment
- Collect information on the treatment process and outcome and related data needed for evaluation of the treatments

The timetable for required follow-up visits will be dictated by various factors, including:

- Requirements for treatment administration and for assessing adherence to treatment
- Rate of occurrence of the outcome(s) of interest
- Patient health care needs
- Cost of a patient visit
- Patient convenience considerations

The schedule for required follow-up visits may be designed to allow for more frequent visits immediately after enrollment of a patient into the trial to permit clinic personnel to initiate and administer the assigned treatment. The interval between visits may be increased to some maximum and held constant thereafter once the initial treatment process is completed.

Follow-up visits that are made on an ad hoc basis because of special problems experienced by the study patients after enrollment into the trial will be variously referred to as unscheduled follow-up visits, nonrequired follow-up visits, or interim follow-up visits.

[1] Patients considered eligible for enrollment into the CDP at the end of the first prerandomization visit were given a single-masked placebo medication (three capsules per day) which they were to take until the randomization visit.

Investigators should construct the data collection schedule so as to be able to distinguish between required and nonrequired follow-up visits. The data system should be designed to yield a count of both types of visits. Differences among the treatment groups in the number of interim follow-up visits can lead to biases in the diagnoses and reports of clinical events used to evaluate the study treatments (see Section 11.6 and Question 68 of Chapter 19 for further discussion).

12.2.4 Visit time limits

Ideally, the entire set of scheduled baseline and follow-up visits for a patient should be done at precise time points relative to the time of randomization. However, such precision is generally not possible in a free-living population, nor is it necessary for most of the observations required in the typical clinical trial. The usual approach is to consider a visit and related data collection as valid if the visit took place within a defined interval on either side of the desired time point. The permissible length of this time window (see Glossary) will depend on the number of required data collection visits and on the amount of variation that can be tolerated in the timing of observations.

The CDP allowed a maximum of 4 months for completion of the three baseline examinations. After enrollment, the patient was required to return to the clinic 1 month after randomization and again at 2 months after randomization for scheduled dosage increases in his assigned medication. Regular follow-up visits were scheduled to take place at 4-month intervals thereafter. Each of these visits had to be within 2 months of the preferred date, as dictated by the date of randomization. Visits not carried out within the time window were counted as missed. The coordinating center for the study provided clinics with computer-generated appointment schedules that indicated the preferred date and the permissible time window for each required follow-up visit (Table 12–1).

12.3 DATA REQUIREMENTS BY TYPE OF VISIT

12.3.1 General considerations

The development of data forms cannot be started until:

- A baseline and follow-up visit schedule has been established by the study investigators
- The purpose(s) of each visit has been outlined
- There is general agreement among the investigators on the specific procedures to be carried out at each visit

A key step in form construction is identification of the specific items of information to be collected during each clinic visit. The process required for the step should be designed to guard against errors of omission as well as errors of commission (see Table 12–2 for a list of precautions). Probably the single most common cause of errors of omission is haste in the development of the data forms. The process of identifying required data items and then constructing and testing them takes time and patience. Efforts to shorten this process in order to get started with patient recruitment and data collection are usually unwise.

The desire to create forms that, in addition to meeting the research aims of the study, provide data needed for routine patient care is probably the single most important contributor to errors of commission. The fact that certain measurements need to be made in providing routine care for patients is not sufficient reason to justify inclusion of them in the study data system.

Before starting form construction, the types of data needed and the procedures for generating them should be outlined. Once developed, the outline should be reviewed by personnel not directly involved in constructing the data forms as a check against the two kinds of errors mentioned above. Further, there should be general agreement on the ordering of the procedures to be performed at any given data collection visit before the forms are constructed. The ordering will influence the sequencing of items on the forms.

One of the last steps in the construction process is to carry out an item-by-item review of each form against a list of data needs and goals, as set down by the leadership of the study. Data items that cannot be justified in this review should be deleted from the final data forms. All follow-up forms should also be checked against each other and against the baseline set of forms for consistency and as a safeguard against errors of omission.

12.3.2 Data needed at baseline visits

The first step in the design of any set of forms is to enumerate the types of data needed (see Section 12.2.2). Baseline data are needed:

Table 12-1 Sample appointment schedule and permissible time windows, as adapted from the Coronary Drug Project

Patient Name: John D. Doe — Patient ID No.: 59-0021

Date of entry: Oct. 31, 1966

Bottle number assigned: 2

The indicated visits should be done within the time windows specified and as close to the desired date as possible. Visits not completed within the sepcified time window should be skipped and will be counted as missed.

Visit	*Desired date*	*First possible date*	*Last possible date*	*Interval length in days*
Dosage adj. Visit 1	Dec. 1,66	Nov. 16,66	Dec. 16,66	31
Dosage adj. Visit 2	Dec. 31,66	Dec. 17,66	Jan. 15,67	30
Follow-up visit 1	Mar. 2,67	Jan. 16,67	May 1,67	106
Follow-up visit 2	July 1,67	May 2,67	Aug. 31,67	122
Follow-up visit 3	Oct. 31,67	Sep. 1,67	Dec. 31,67	122
Follow-up visit 4	Mar. 2,68	Jan. 1,68	May 1,68	122
Follow-up visit 5	July 1,68	May 2,68	Aug. 31,68	122
Follow-up visit 6	Oct. 31,68	Sep. 1,68	Dec. 31,68	122
Follow-up visit 7	Mar. 2,69	Jan. 1,69	May 1,69	121
Follow-up visit 8	July 1,69	May 2,69	Aug. 31,69	122
Follow-up visit 9	Oct. 31,69	Sep. 1,69	Dec. 31,69	122
Follow-up visit 10	Mar. 2,70	Jan. 1,70	May 1,70	121
Follow-up visit 11	July 1,70	May 2,70	Aug. 31,70	122
Follow-up visit 12	Oct. 31,70	Sep. 1,70	Dec. 31,70	122
Follow-up visit 13	Mar. 2,71	Jan. 1,71	May 1,71	121
Follow-up visit 14	July 1,71	May 2,71	Aug. 31,71	122
Follow-up visit 15	Oct. 31,71	Sep. 1,71	Dec. 31,71	122

Source: Reference citation 104. Adapted with permission of the American Heart Association, Inc., Dallas, Texas.

- To establish patient eligibility through items that indicate the presence of required eligibility conditions and the absence of exclusion conditions
- To characterize the demographic and general health characteristics of patients eligible for enrollment into the trial
- To establish a baseline for assessment of changes in variables to be measured over the course of follow-up
- For any stratification required in the randomization process
- For post-stratification
- To aid in contacting and tracing patients
- To assess clinic performance in carrying out the informed consent process
- To assess adherence to the study protocol
- To link baseline and follow-up records
- To address other topics unique to the study in question

The second step is to list the specific data items and forms needed for each visit. Some items will appear only once in the list; others will appear under several categories.

The need for record linkage can usually be satisfied by use of a unique number that identifies the patient and type of visit performed. The data needed for stratification will be satisfied by collection of information necessary for making the classifications called for in the stratification. Variables that are to be tracked over time must be observed during the prerandomization or randomization visit to provide the necessary baseline information. The same is true for variables that are to be used in risk-factor or subgroup analyses to be carried out later on in the trial. Investigators must have a thorough knowledge of the epidemiology of the disease being treated and of the conditions likely to influence the selected outcome measures to make an intelligent choice of baseline variables for use in such analyses.

Table 12-2 Methods for avoiding errors of omission and commission in the data form construction process

A. Safeguards against errors of omission

- Allow adequate time for developing and testing data forms before starting data collection
- Solicit content advice and input from persons not directly involved in the development process
- Review data forms used in similar trials
- Ask persons not directly involved in the developmental process to review proposed data forms for deficiencies
- Test data forms under actual study conditions before use in the study

B. Safeguards against errors of commission

- Distinguish between data needed for patient care and those needed to address the objectives of the trial
- Make certain every data item scheduled for collection is of direct relevance to achieving a stated aim or objective of the trial
- Establish an appropriate set of review and approval procedures in order for new items to be added to existing data forms

12.3.3 Data needed at follow-up visits

Data collected during follow-up are needed to:

- Assess changes in variables that are or may be affected by treatment
- Characterize the nature of treatment over the course of follow-up
- Characterize departures from the treatment protocol and the reasons for them
- Characterize patient adherence to the assigned treatment(s)
- Characterize the nature of treatment effects observed, including side effects and patient complaints related to treatment or believed to be related to treatment
- Characterize the state of a patient's health and quality of life
- Maintain up-to-date patient locator information
- Assess adherence of clinic staff to required procedures, as set down in the study protocol
- Link baseline and follow-up records obtained on the same patient
- Address other topics unique to the study in question

The same process as outlined for baseline forms should be used to construct the follow-up forms. It should begin with an enumeration of items related to the above categories. It is wise to identify all the variables on the baseline set of forms that are to be updated at one or more follow-up visits before starting construction of the follow-up forms. Once this is done, it is necessary to indicate the visit or visits at which specified variables are to be observed.

A series of items will be required to provide data on treatment administration. Trials involving technically complicated treatment procedures, such as in some surgical trials, may require an entire set of forms for characterizing the treatment process.

The follow-up data system must also provide information on treatment compliance and on the amount of exposure a patient has had to competing treatments. The latter information is needed to characterize the extent of cross-treatment contamination present in the various treatment groups when the results of the trial are analyzed. The follow-up forms must also include items for recording real or imagined treatment side effects reported by the study patients. A thorough knowledge of the treatments being tested and of pertinent medical literature is needed to formulate suitable items.

A category of major interest in some trials (e.g., cancer chemotherapy trials) concerns the effect of treatment on a patient's quality of life. The outcome measure, whether it be death or some nonfatal clinical event, may be only part of what is needed for treatment assessment. A test treatment, even if known to prolong life, may be rejected by patients because of its noxious side effects. Information on changes in a patient's employment status, recreational activities, exercise habits, ability to care for himself, etc., will be needed if quality of life measures are to be used in evaluating the study treatments.

12.4 CONSIDERATIONS AFFECTING ITEM CONSTRUCTION

12.4.1 Implicit versus explicit item form

A key consideration in item construction has to do with wording of the items and whether they are stated in explicit or implicit terms. Examples of the two forms are given below.

Explicit item form

What is your present age? __ __ *Age in Years*

What is your birthdate? __ __ - __ __ - __ __ *Mo Day Yr*

Implicit item form

Age . . . . __ __

Birthdate __ __ - __ __ - __ __ *Mo Day Yr*

The wording chosen will depend upon the nature of the information being collected and on the level of sophistication of the person responsible for completing the items. An explicit form is needed when the wording of an item can effect the information to be obtained. Survey researchers have long recognized the importance of standardized wording for questions when the information is collected via an interview.

An implicit form may be satisfactory for items completed by clinic personnel. However, even in this case, care must be taken to make certain the item is constructed so as to avoid misinterpretation among staff responsible for completing the item.

12.4.2 Interviewer-completed versus patient-completed items

The data forms may be designed to be completed by clinic staff or by the patients themselves. Most of the forms will be completed by clinic personnel in a clinical trial. Hence, the remainder of this chapter and Appendix F is written from this point of view. However, many of the same points outlined in Sections 12.5 and 12.6 apply to forms completed by patients as well.

Items used as a reminder to clinic personnel to obtain certain information should be distinguished from those that are to be read or presented to the patient exactly as they appear on the form. The Hypertension Prevention Trial (HPT) preceded all items of the latter type by letter codes of AAW—Ask-as-Written—or SAW—Show-as-Written (see examples below). Items that had a long list of possible answers or were considered too complicated to comprehend via a verbal presentation were presented in the SAW fashion using specially prepared flashcards. The participant selected his response from among those listed on the card, either by pointing to the proper line on the card or by reading his reply from the card.

Example of Ask-as-Written item

(AAW) Are you presently taking vitamins or minerals regularly?

( ) ( )
Yes No

Example of Show-as-Written item

(SAW) Have you taken any of the following drugs in the last month? (Use HPT Flashcard 04 and check as many as apply)

( ) Anacin
( ) Appedrine
( ) Bromoquinine
( ) Coryban D
( ) Dexatrim
( ) Dristan
( ) Excedrin
( ) Midol
( ) Nodoz
( ) Permathene-12
( ) Prolamine
( ) Triaminicin
( ) Vanquish

The SAW approach can be useful in the collection of sensitive information involving personal income, sexual behavior, or the like. A patient may be more willing to indicate his reply by pointing to the appropriate reply or by referring to a letter or number code on a flashcard than to answer the question verbally. Other techniques have been developed for collection of sensitive information. A particularly interesting one involves a "random response" technique. The technique is not discussed herein, but descriptions and illustrations of it can be found in papers by Bégin et al. (1979), Bégin and Boivin (1980), Frenette and Bégin (1979), Himmelfarb and Edgell (1980), Martin and Newman (1982), and Zdep et al. (1979).

12.4.3 Questioning strategy

The designers of the data forms must decide where general, nondirective, questions are to be used to elicit subjective information and where more specific, directive ones are to be used. Clearly, the type and amount of information obtained can be influenced by the questioning strategy used. For example, the number of patients reporting gastrointestinal distress in an aspirin study can be expected to be higher if the count is based on responses to a specific question concerning such problems (e.g., Have you had any gastrointestinal distress since you started treatment?), as opposed to a general question (e.g., Have you had any problems since you started treatment?). The two strategies may be used in tandem in situations in which it is appropriate to begin an area of inquiry with a general question followed by one or more that are specific and direct.

12.4.4 Single versus multiple-use forms

The organization and content of the forms will be influenced by whether they are designed to be completed over a series of clinic visits or at a single visit. Multivisit forms are more efficient to use in that there are fewer forms to complete and process than is the case with single-visit forms. Further, since there is some administrative overhead associated with the completion and processing of any form, the fewer the forms, the lower the total overhead.

A disadvantage with multivisit forms is the time and inconvenience involved in filing and retrieving partially completed forms. Further, their use can slow the flow of information in the study since a form cannot be sent to the data center until it is complete. The delay can be lengthy if the visits to be covered are widely separated in time. Hence, if they are used at all, their use should be limited to sets of visits that are completed over short time intervals.

Data generated at different sites, whether within or outside the clinic, even if part of the same visit, should be recorded on separate forms. This is particularly true for forms used to record results of procedures or measurements that are done by personnel who are not under the direct control of the study clinic and that cannot be provided on the day of the patient's visit to the clinic (e.g., as is usually the case with most laboratory determinations and with expert readings of biopsy materials, coronary angiograms, ECGs, eye fundus photographs, and the like). The only exceptions are those in which the data in question flow back to the study clinic within a day or two of the patient's clinic visit.

12.4.5 Format and layout

Decisions need to be made regarding the general format and layout of the data forms. Issues to be addressed include (see Section 12.6 for discussion):

- Full-page versus multicolumn layout
- Paper size, quality, and color
- Use of boxes, parentheses, or lines for recording responses to designated items
- Location of check spaces for responses
- Printed versus photocopied forms

12.5 ITEM CONSTRUCTION

This section and the next contain a series of detailed comments and suggestions concerning item and form construction. Many of the points are supported with illustrations contained in Appendix F.

12.5.1 General

1. Every item and item subpart should have a unique identifying number (Appendix F.1).
2. Items should always be constructed to require a response, regardless of whether a condition is present or absent. The practice of allowing a blank or unanswered item to indicate the absence of a condition can cause confusion. Once the form is completed there is no way to distinguish between items purposely left blank because the condition in question was not present from those accidentally left blank (Appendix F.2).
3. The conditions under which an item is to be skipped should be part of the item or should be included in the instructions for the item (Appendix F.2.4).
4. Items or sections on a form that may be skipped in certain instances should be preceded by items that document the legitimacy of the skip. For example, a form should include an item for recording the patient's age if parts of the form are to be skipped for patients in a specific age range.

12.5.2 Language and terminology

5. Use simple, uncomplicated language.
6. Avoid the use of esoteric terms and abbreviations. This is especially important in situations where there is likely to be a turnover in the personnel responsible for completion of the study forms, or in multicenter trials where the level of staff familiarity with the study forms may vary.
7. Avoid the use of terms that may have different meanings to the different people involved in completing the forms.
8. Provide necessary definitions on the forms or indicate where they may be found.
9. Use simple sentences in the construction of items and instructional materials. Phraseology should be consistent with the educational level of the individuals responsible for completion of the forms.
10. Avoid unnecessary words (Appendix F.3).
11. Avoid the use of double negatives (Appendix F.4).
12. Avoid the use of compound questions by dividing them into a series of specific questions (Appendix F.5).
13. Items requiring a comparative judgment should indicate the basis for the comparison (Appendix F.6).
14. Language research suggests that positive terms, such as better, bigger, or more, are less subject to interpretation error than negative terms, such as worse, small, or less (Wright and Haybittle, 1979a) (Appendix F.6.3).
15. Items requiring an affirmative or negative response are confusing when an affirmative reply indicates the absence of a condition (Appendix F.7).
16. For the same reason as indicated in 15, questions concerning disease state or history are easier to understand if stated in a way which requires a yes or positive reply when the condition is present, rather than when it is absent (Appendix F.8).
17. The time point or interval to be used in answering an item should be explicitly stated in the item. A time point may be defined by a specified date, by some event or condition, or simply as the "present." A time interval may be defined by two calendar dates or from some date to the present (Appendix F.9).
18. Variation in the direction of response from question to question (e.g., stating some questions that require a comparative assessment in positive terms and others in negative terms) should be avoided (Appendix F.10).
19. Avoid leading questions (Appendix F.11).

12.5.3 Use of items from other studies

The item construction process can be facilitated by a review of existing forms from related studies. The review may help identify data items that should be included on the data forms as well as aid in their construction and format.

20. Assemble sets of forms from other related studies and order by topic (e.g., smoking history, exercise habits, disease history, and so on).
21. Do not use an item *simply* because it has been used before in other studies.
22. Do not construct an item de novo if a suitable version of the item already exists, has been used in other studies, and has seemingly produced reliable information.
23. Do not modify the wording of an item taken from another study if the item has been shown to produce useful information and if information generated from it is to be compared with findings from studies in which the item was used.
24. Do not use an entire form or section of a form that has been copyrighted without the written approval of the copyright holder.
25. Do not reproduce an entire form or section of a form used in another study without permission from the study, even if the form is not copyrighted.

12.5.4 Closed- versus open-form items

A closed-form item is one that is completed using a defined list of permissible responses. An open-form item is characterized by the absence of a defined list of permissible response options.

Closed-form examples

Indicate the highest grade completed in school:

( ) 6th grade or less
( ) 7th, 8th, or 9th grade
( ) 10th or 11th grade
( ) 12th grade
( ) 2 or 3 years of college
( ) 4 years of college
( ) 5 or more years of college

Have you had any of the following diseases or conditions diagnosed in the last year? (check all that apply)

( ) Heart attack
( ) Stroke
( ) Congestive heart failure
( ) Emphysema
( ) Cancer
( ) None of the above

Open-form examples

What is the highest grade you have completed in school?

____

Use the space below to list serious illnesses that you have had. (Enter "none" if you have never had a serious illness.)

____

____

26. An open-form item should be used when it is difficult to anticipate the different responses that may be given, or when there is a desire to avoid leading the respondent by indicating permissible replies.
27. An open-form item should be used to record continuous data, unless a closed form, with designated categories, is considered to provide adequate detail (see Section 12.4.2). An open form should be used even if data are to be subsequently tabulated into designated categories (e.g., age <25, 25–49, and ≥50). The opportunity to categorize in different ways is lost whenever continuous data are collected and recorded in categorical form.
28. Closed-form items, with a predefined list of response options, should be used when there is a need to structure the responses obtained (e.g., when it is desired to present the respondent with all possible options when answering a question or when it is desirable to remind him of the permissible response options).
29. The time required to code and process information from open-form items is usually greater than for closed-form items.
30. A closed-form item will do little to facilitate coding and processing if most of the responses fall into a general catchall category, such as the "other (specify)" category, included at the end of the response list.

12.5.5 Response checklist

A response checklist defines the permissible or acceptable responses to an item. The simplest checklist is one for items requiring a binary response, such as yes or no, present or absent, or the like. This list should cover all possible responses and may be constructed to allow only one response or multiple responses, depending on the item.

31. A response checklist is preferable to an unformatted written reply, except as indicated in Section 12.5.4. An item involving a long list of possible response options (see Section 12.4.2 for flashcard alternatives) will require more space for layout than an item designed to elicit an unformatted written reply, but the information generated will be easier to process and interpret than is the case with an unformatted written reply.
32. Vertical checklists are easier to use and are subject to less confusion with regard to the location of appropriate check spaces than are horizontal checklists (Appendix F.12).
33. A response checklist that is not exhaustive should include an "other" category that can be used to record responses not covered in the list.
34. There should be adequate space on the form for respondents to write out responses that fall into the catchall category. The space provided will influence the amount and legibility of the information recorded.
35. Frequent use of a catchall category for an item increases the time required for completion of the item and for coding and processing the information generated by it (assuming the written responses are to be coded and processed).
36. It may be wise or necessary to expand the list of permissible response options for an item during the trial. Any expansion should be based on a review of the responses provided in the catchall category and should be done as soon after the start of data collection as is feasible. Expansion may not be practical in short-term trials or in situations in which it can be expected to cause major coding or analysis problems.
37. A condition is more likely to be recorded as present if it appears in a checklist than if it does not. Hence, list expansions during the trial may appear to "increase" the prevalence of certain conditions. However, the expansion will not influence treatment comparisons unless the changes were implemented at different times for the various treatment groups under study.
38. It is sometimes convenient to include a summary check position at the head or end of a list that may be used in lieu of checking each individual entry for the list (see Appendix F.12.2.4 for example).

12.5.6 *Unknown, don't know,* and *uncertain* as response options

39. The three options are interrelated and are to a large extent used as if they were interchangeable. The particular option listed will depend on the context of the question.
40. The operational implications are about the same. All three options imply the lack of information needed to answer a question.
41. *Don't know* or *uncertain* should not be listed as a response option if the aim of the item is to require the respondent to record his best guess even if he does not know or is uncertain regarding the accuracy of his reply. The form should have written instructions when guesses are required.

12.5.7 Measurement and Calculation items

A measurement item is one that requires the respondent to record some measurement. A calculation item is one that requires the respondent to carry out an arithmetic calculation using other information on the form. The examples that follow are taken from the HPT.

Height and weight measurement and calculation example

Height (shoes off): __ __ *inches*

Weight (outdoor garments and shoes off): __ __ __ *lbs*

Q.I. = Wt/Ht² 0. __ __ __ __ *lbs/in²*

Blood pressure measurement and calculation example

	BP in mm Hg	
	SBP	*DBP*
1st RZ BP		
a Reading	__ __ __	__ __ __
b Zero value	__ __	__ __
c a−b	__ __ __	__ __ __
2nd RZ BP		
d Reading	__ __ __	__ __ __
e Zero value	__ __	__ __
f d−e	__ __ __	__ __ __
Average RZ BP		
g Sum (c + f)	__ __ __	__ __ __
h Avg (g ÷ 2)	__ __ __	__ __ __

42. The unit of measurement should be specified on the form (Appendix F.13).
43. Measurements should be made and recorded in units familiar to the personnel responsible for making them. Use of an unconventional unit may lead to data collection and recording errors (Appendix F.13.4)
44. Whenever feasible, all recordings of a specified variable should be made using the same unit. Use of different units may occur when different laboratories are used (e.g., as in a multicenter trial in which each clinic relies on its own laboratory for making required laboratory determinations).
45. Space should be provided on the form for the respondent to indicate the unit of measurement when it is not practical to specify the unit in advance (Appendix F.13.2.2, F.13.2.3).
46. Continuous variables, such as age, blood pressure, laboratory values, and the like, should not be recorded in categorical form (see statement 27).
47. The precision required for a measurement should be specified on the data form (Appendix F.14).
48. The amount of precision required for a measurement should not exceed the error involved in making the measurement (see comment regarding item F.14.1 in Appendix F).
49. The raw data used to make any summary calculations should be recorded on the form.
50. Data forms should be constructed to minimize the number of arithmetic calculations required during a patient visit. All calculations except those needed to perform a patient examination or to carry out some other treatment or data collection function during the examination should be performed at the data center as part of the data entry and analysis processes.
51. Calculations needed on data forms made by clinic staff, even relatively simple ones, should be made using a pocket calculator or a computer.
52. Items requiring a series of arithmetic operations during completion of a data form should be arranged in a format that facilitates those operations. For example, numbers that must be added or subtracted should be arranged vertically and with adequate space for recording intermediate calculations (Appendix F.15).

53. Arrange the calculations for a given item in a single unbroken column, if possible. Avoid arrangements in which calculations are started on one column of a form and continued on the next column or page.

12.5.8 Instruction items

An instruction item is one that is included on a form to instruct the individual completing the form as to how to deal with a given question. The two types of instruction items discussed are STOP and SKIP items (Appendix F.16).

54. A STOP item is used to indicate conditions that, when encountered during the course of a patient visit, require clinic personnel to temporarily or permanently halt some procedure or process. The stop will be permanent unless the conditions that require the stop can be removed.
55. STOP items on any given form should be arranged to allow the respondent to terminate all work on the form as soon as a stop is checked. This requires an arrangement in which essential information, required on all patients, is obtained before any stops are allowed.
56. A common use of STOP items during the prerandomization series of clinic visits is to indicate conditions that exclude a patient from enrollment into the trial. Stops of this sort will halt further work-up of the patient.
57. It is wise to arrange prerandomization stops for procedures in ascending order with regard to the risk or general discomfort they entail for patients. The goal should be to carry out the lowest risk, least expensive, most productive procedures first.
58. A SKIP item may be used whenever there is an item or series of items on a form that can be skipped depending on the answer to the item.
59. A SKIP item should indicate the conditions under which the skip can occur and the item or items to be skipped.

12.5.9 Time and date items

A time item is one that requires the respondent to record the actual clock time at which some step, procedure, or measurement was carried out. A date item is one that indicates the date some step, procedure, or measurement was carried out (see items in Section F.13.1 of Appendix F for examples).

60. Items requiring a clock time should indicate whether the time recordings are for A.M. or P.M. if a 12-hour recording system is used. The use of A.M. and P.M. will cause confusion for recording 12 noon and 12 midnight, unless instructions given on the form indicate how these times are to be recorded.
61. Times should not be recorded on a 24-hour basis unless personnel responsible for the recordings are thoroughly familiar with 24-hour timing schemes or the readings are made directly from 24-hour clocks.
62. The order to be used in recording the date should be specified on the form (e.g., see items in Section F.13.1 in Appendix F). The two most common conventions are

    Month, Day, Year
    Day, Month, Year

63. Failure to specify the convention to be used dates if they are recorded in digital form. For example 1-9-82, could be read as January 9, 1982, or 1 September 1982, depending on the convention used.

12.5.10 Birthdate and age items

64. The baseline data forms should include both an age and birthdate item if age is to be used either as an eligibility condition for enrollment into the trial or in subsequent data analyses.
65. Date of birth is a key piece of information in many trials. It may be needed for making accurate age calculations or for record search and linkage operations in the follow-up of dropouts for mortality via the National Death Index and other similar files. Birthdate may also be useful in linking different records for the same individual if name is not collected.
66. A patient's reported age should be checked against his reported birthdate on entry into the trial, as illustrated in Appendix F.17. This is particularly important if age is used as an eligibility condition. Discrepancies should be resolved.
67. The age that is reported may differ depending on the source it is taken from. For example, insurance companies consider a person to have attained the next year of age one-half year beyond his last birthday anniversary, whereas a person reporting his age will give it as of his last birthday anniversary.

12.5.11 Identifying items

Every data form should contain space for recording the patient's ID number and name (or name code). Once these two items have been entered into the data system, a cross-check should be made on all new data to be entered. Information from a form should not be added to the data system if the ID number and name (or name code) do not agree. See Section 12.6.9 for further comments.

68. It is wise to construct a name code, made up of some combination of letters from the patient's first, middle, and last name, for use as a patient identifier. This identifier is in addition to ID number and should not be changed once it has been issued, even if the patient has a subsequent name change. The name code may be used in addition to name or in place of it depending on whether the study forms are designed to preclude collection of name.
69. Each follow-up data form should include an item for recording visit number. The number is typically checked against the patient's appointment schedule (see Table 12-1 for example) to determine if the visit occurred within the permissible time window.
70. Patient identifiers useful for mortality follow-up include:
    - Social Security number
    - Date of birth
    - Place of birth
    - Father's name
    - Mother's maiden name
    - Patient's maiden name for females
    - Date and place of death (if applicable)
71. A unique identifier should be assigned to each member of the clinic staff involved in data collection. This number may be used in place of name or initials (or in combination with name) to identify the individual responsible for completing or reviewing a form, or a series of items on a form.

12.5.12 Tracer items

A tracer item is one that is used to obtain information needed to locate a patient. In some cases the information provided by such items is used to locate and recontact a patient who has dropped out of the study to try to persuade him to return to the clinic for examination and subsequent follow-up. In other instances the items are used to facilitate the collection of mortality or morbidity data.

72. Tracer data should be collected on all patients upon entry into the trial and should be updated at periodic intervals over the course of the trial.
73. Useful patient tracer data include:
    - Current address and telephone number (home and work if patient has both)
    - Employer's name, address, and telephone number
    - Name, address, and telephone number of a close relative
    - Name, address, and telephone number of a friend or neighbor
    - Name and address of patient's private physician
74. Other tracer items, especially for mortality follow-up, are listed in Section 12.5.11, Statement 70.

12.5.13 Reminder and documentation items

A reminder item is one that is intended to remind clinic personnel to perform an indicated procedure or task (Appendix F.18.1). A documentation item is one that is used to indicate that a step or condition required in the data collection, enrollment, treatment, or follow-up process has been performed (Appendix F.18.2).

75. Reminder items are useful in trials with complicated data collection schemes, or in which there is a good chance that some of the personnel involved in data collection will be unfamiliar with details of the data collection protocol.
76. Reminder items should be used in conjunction with steps or procedures that are essential to the data collection, enrollment, treatment, and follow-up processes.
77. Key data items that are to be completed by a designated individual should be followed by documentation items for recording the

date the items were completed and the name or certification number of the individual who was responsible for their completion.

78. There should be space at the end of each form to record the date the form was completed.
79. Documentation items should be included at the end of each form for recording the name or certification number of the person responsible for review of information on the form and for recording the date of the review (Appendix F.18.2).

12.6 LAYOUT AND FORMAT CONSIDERATIONS

12.6.1 Page layout

80. Choose a layout that permits use of a single page size for all forms (e.g., 8½" × 11").
81. Use a layout in which all pages within a form are oriented in the same way. That is, with pages laid out either portrait style (i.e., with lines of print running across the short axis of the page) or landscape style (i.e., with lines running across the long axis of the page).
82. If possible, use the same page orientation for all forms of a given type (e.g., all those used at the clinic for follow-up data collection).
83. Use a layout that is uncluttered and that facilitates use of the forms by both clinic and data processing personnel.
84. Choose between a full page or two-column layout (Appendix F.19).
85. Generally, two-column layouts are more space efficient than full page layouts.
86. The layout chosen should be compatible with the data entry needs of the study; clinic needs should take precedence over those for data entry if meeting both needs leads to conflicting layout requirements.
87. Avoid a layout such as that displayed in Appendix F.19.1.1, where check spaces are scattered over the page. The layout increases the time required to complete and key a form and may contribute to errors in those processes as well.
88. Use layouts such as those illustrated in Appendix F.19.1.2 and F.19.2. Standardizing the location of check positions within and across forms facilitates completion of the forms and reduces the time and errors involved in keying data from them.
89. Whenever feasible, choose a layout that facilitates entry of data directly from the form, such as illustrated in Appendix F.19.2.
90. Items should be arranged so as to minimize the number that are split across columns or pages of a form.
91. The pages of a form should be printed or typed on only one side. The reverse side of the pages may be used to print instructional material or should be left blank.
92. Page layout should be designed to help respondents identify items or sections of a form that are to be skipped under specified conditions. This may be done by setting key words or phrases in boldface type or by use of special instructions or other aids to direct the respondent to applicable items or sections (Appendix F.20).
93. The space between subparts of an item should be less than the space between items.
94. The space separating items should be uniform unless variation in spacing has operational significance.
95. Similarly, the space separating one part or section of a form from another should be the same and should be greater than the space separating individual items.
96. Right-hand justification of typed or printed text should be avoided if it results in noticeable variation in the spacing between words.

12.6.2 Paper size and weight

97. Use a good quality paper with enough gloss to avoid bleeding through from ink or felt pens.
98. Use the same size paper for all forms (see statements 80, 81, and 82).
99. A paper size of 8½" × 11" is preferable to other sizes, especially when forms are to be photocopied and filed using standard office equipment.

12.6.3 Type style and form reproduction

100. The print or type font used should be large and crisp enough to allow for image degradation when forms are photocopied.
101. Use a print or type font at least the size of newsprint.

102. Avoid capitalization of long phrases or sentences. Text written in capital letters is more difficult to read than a mixture of upper- and lower-case letters (Wright and Haybittle, 1979b).
103. Use a different print or type font for emphasizing specific words, phrases, and headings and for distinguishing instructional material from data collection items (e.g., see items F.21.1 in Appendix F).
104. Printed forms are generally easier to read and are esthetically more pleasing than typewritten forms.
105. Consideration should be given to printing forms that are to be used in large numbers or that are difficult to photocopy because of their size or the way in which they are assembled. Forms should not be printed until they have been thoroughly tested and are no longer subject to revision. It may be less costly to photocopy forms that are used in small numbers. The same may be true for forms used in relatively large numbers if they are likely to undergo changes. Forms may be photo-reproduced from either typed or professionally printed masters.

12.6.4 Location of instructional material

106. Instructional material on the first page of the form should indicate when the form is to be used and who is responsible for completing it (Appendix F.21.2.1).
107. Instructional material relating to specific items or sections of a form should be located next to those items or sections (Appendix F.21.2.2).
108. All instructions needed for completion of a form should be included on the form. This is especially important in long-term trials in which personnel may change over the course of the trial, and in multicenter trials.
109. All instructional material should be as concise and simple as possible.
110. Instructional material should be identified by use of a special type font or in some other way (Appendix F.21).
111. Instructional material that is too extensive for inclusion next to the item or section to which it pertains should be contained in a separate booklet or should appear on the back side of the page adjacent to the one in question.
112. Key definitions needed for completion of an item should appear on the form.
113. The instructions should identify items that are to be read verbatim to the patient, as discussed in Section 12.4.2.
114. Items with a list of permissible responses that are not mutually exclusive should contain an instruction to indicate whether or not the respondent may check more than one response.
115. Items which include *unknown, don't know*, or *uncertain* as response options should include instructional notes to indicate if any special procedures are required before these categories are checked (e.g., an instruction to remind clinic staff to check specific medical records before checking the uncertain category for a designated item).
116. The instructions should indicate the steps to be followed in performing a particular measurement or procedure. Reference to the appropriate section of the study handbook or the manual of operations should appear on the form if the measurement or procedure is too complicated to be outlined on the form.
117. There should be an instruction at the end of each form that indicates where the form is to be sent after completion and the steps to be followed in preparing the form for transmission.

12.6.5 Form color coding

Color coding is useful if there is a need to distinguish among different types of forms (e.g., prerandomization forms versus follow-up forms, or forms completed in the laboratory versus those completed in the clinic) or among different copies of the same form (e.g., white for the original, green for the first copy, and pink for the second copy).

118. The color-coding scheme should be simple, logical, and easy to remember.
119. The colors chosen should be limited to a few distinct shades.
120. A particular color should have the same meaning throughout the study (e.g., pink always identifies the second copy of an original).
121. As a rule, forms printed on pastel-colored paper are easier to read and will produce better quality photocopies than those

printed on dark-colored paper. The legibility of photocopies produced from pages using the colors proposed should be checked before making the final color selection.

122. Color coding should never be used as the sole means of identifying a form or its use. Written information should appear on the form to designate its use and should be sufficient to identify a particular form if individuals are unable to distinguish among the colors.
123. It may not be practical to use multicolor forms if a clinic is responsible for maintaining its own supply of forms from photocopy masters.

12.6.6 Form assembly

124. Multipage forms may be supplied to clinics collated and bound (e.g., stapled), collated and unbound, or uncollated. The latter method of supply is preferable when the number of pages making up a form varies depending on the patient or examination. Forms that are collated should be supplied unbound if it is likely that they will have to be disassembled for completion or to make photocopies of them after completion.
125. The individual pages of a form should be sequentially numbered and should indicate the total number of pages in the form (e.g., by using the following kind of numbering scheme: page 1 of 10, page 2 of 10, etc.).
126. Paper clips or similar kinds of fasteners are not acceptable for securing the pages of completed forms. They are likely to come off as the forms are handled in copying, coding, or filing.
127. Forms may be developed with specially designed answer pages that may be detached from the main body of the form. The Lipid Research Clinics used this approach to reduce the volume of paper flowing to the coordinating center. Detachable answer pages may be used only if all information required for data entry can be recorded on the answer sheets and adequate documentation is provided on the answer sheet to identify the patient and type of examination performed.

12.6.7 Arrangement of items on forms

Thought should be given to the ordering of items within and across forms. The arrangement should be compatible with the needs of patients and clinic staff. Arrangements that are not may result in missed or poor quality data.

128. Place items calling for a particular frame of reference next to one another.
129. The nature, quality, and quantity of information obtained on a form may be influenced by the order of the items on it.
130. The number of positive responses to a list of questions will be higher for lists that are read or shown to the patient than when the list is simply used by clinic staff to record information volunteered by the patient (see Section 12.4.3).
131. The order of procedures should remain fixed over the duration of the trial, especially if there is any chance that one procedure (e.g., ingestion of iopanic acid in order to perform cholecystograms) affects the results of another procedure (e.g., serum cholesterol determinations; see National Cooperative Gallstone Study Group, 1981a, for additional details). A fixed order does not necessarily eliminate this problem, but it does control the effect over time and across treatment groups. Further, not all variations in sequencing can be avoided if the number of procedures performed differs from examination to examination.
132. The arrangement of items within a form should be compatible with the preparation required for a particular examination (e.g., the items to be completed with the patient in a fasting state should appear before those that are to be completed after the patient has been allowed to eat or has been given a glucose load).
133. Group items into sections with headings indicating the general content of the sections. Use a different type font to facilitate identification of section headings.
134. The numbering and identification scheme used on a form should be designed to facilitate the identification of items and their subparts.
135. Use different spacing to indicate transitions from one item to another and from one section to another.
136. Devise a numbering system for identification of individual items on a form. Items should be numbered sequentially over the entire form or within sections of the form. The former system is preferable. The latter one has the advantage of allowing for addition or deletion of items in a section without disrupting the numbering system for other sections. However, the disadvantage is that both a section and item number are needed to locate a specific item on a form.
137. Items should be arranged among forms so that any given form can be completed in a single session, as discussed in Section 12.4.4.
138. The time lag between collection of a block of information and transmission of that information to the data center should be minimized. This generally requires use of different forms for recording data that are generated at different clinic visits. Different forms may be needed as well for data generated at the same visit, but by people at different locations in the clinics.
139. Data items that are considered confidential or that deal with sensitive information should appear on separate pages of a form or on a different form so that it is possible for the page or form to be stored apart from the remainder of the patient's file.

12.6.8 Format

12.6.8.1 Items designed for unformatted written replies

Items in this class should provide space for handwritten replies without any restriction on the number of characters of information that may be provided (Appendix F.22).

140. The amount of space provided on the form will influence the quantity and quality of information supplied.
141. The space provided should be consistent with the amount of detail desired and should be large enough to prevent the respondent from having to resort to use of cryptic abbreviations or unnaturally small handwriting.
142. Designate the area where the reply is to be recorded. If lines are used, the space between them should be at least ¼″ (e.g., see item F.22.2 in Appendix F).
143. An unlined space, such as shown in item F.22.3, may be preferable to use of lines, especially if responses are typed.

12.6.8.2 Items requiring formatted written replies

Items in this class require the respondent to fit the response into a designated number of character spaces. The restriction is ordinarily imposed to facilitate processing of the information.

144. The number of allowable characters per item will be dictated by the code format established when the item was developed.
145. Formatted items should indicate the number of data characters allowed or required. This may be done in the instructions accompanying such items (e.g., by asking the respondent to make certain his reply does not exceed more than a specified number of characters) or by using character boxes or lines, as illustrated in Appendix F.13.1 and F.23.
146. Character lines are preferable to character boxes, especially if the lines that form the boxes serve to camouflage characters contained in the boxes. The weight of the lines or color of the ink used to form the boxes should be distinctly different from the line weight or color of the characters appearing in the boxes when boxes are used.
147. Forms to be completed by hand should have character line segments that are ≥¼″ long. The line segments may be shorter if the forms are to be completed using a typewriter.
148. The precision requirements for numeric data should be indicated in the item, as illustrated in Appendix F.14.1 or F.14.2.

12.6.8.3 Items answered by check marks

149. The order of responses (e.g., yes followed by no, or vice versa) should be uniform throughout a form and across forms (Appendix F.24).
150. Inadequate space for checking the proper response (Appendix F.24.4) may lead to errors when items are completed or keyed. The separation of check spaces when arranged vertically may have to be fairly sizable if multiple copies of a form are to be made using carbon or NCR (no carbon required) paper. Variation in the registry of the copies relative to the master can render entries recorded on the copies ambiguous.
151. The space used for checking a response should be as near the items as possible. A dashed or dotted line should be used to associate the check space with the response category when the latter is widely separated from the former (see Appendix F.24.8 and F.24.9).
152. A long list of response options should be broken by a blank line after every third or

fourth entry in the list to aid the eye in locating the appropriate check space (Appendix F.24.9 and F.24.10).

153. Forms requiring a check mark to indicate the appropriate reply to a question are preferable to those in which the respondent reads a list of items associated with the question and then records the code number(s) of the item(s) selected. The latter approach should be considered only when the same list of responses applies to several different questions on the form, or when the list of possible responses is inordinately long.
154. Use of lists that are not part of a form, or that are located elsewhere on it, may increase the time needed to complete the form.

12.6.9 Location of form and patient identifiers

155. Each form should bear the name of the study, the name of the form, a form number, version number, and version date.
156. The form number, version number, and version date should appear on each page of the form. The version date is useful if individual pages are revised during the study.
157. There should be space on each page for recording the patient ID number and visit number (see Section 12.5.11).
158. The space for recording patient ID number should appear in the same relative position on all forms (e.g., upper right-hand corner). A standard location helps to minimize the risk of the item being left blank when forms are completed and facilitates use of the information for filing and retrieval.

12.6.10 Format considerations for data entry

159. If possible, data forms should be designed to allow for data entry directly from the form, without intervening transcription of the data. This generally requires designation of codes and fields on the form (Appendix F.25), except where data entry is done via CRT screens that display the required fields.
160. It may be useful to reserve space on each form for office use. The space may be used to record transactions involved in the completion of the form and entry of information into the data system.
161. Coding and data entry operations should be designed to minimize the number of times a form is handled. Ideally, all information should be keyed at the same time including any handwritten unformatted information.
162. A special code should be entered into the data system to identify items that contain data that are not keyed (e.g., uncoded handwritten replies). The code is useful if it is ever necessary to retrieve forms containing unkeyed information.
163. The location of check spaces should be standardized to facilitate the data entry process.
164. Coding conventions should be uniform across forms (e.g., use the same letter or number code to denote a yes reply).
165. The layout of a form should take account of coding and data entry requirements, but should not be dominated by them, especially if the layout complicates use of the form in the clinic.
166. The coding layout should permit data entry personnel to proceed through a form in an orderly fashion with few, if any, references to items already keyed or to items still to be keyed.
167. The form number, version number, or version date appearing on a completed form should be keyed. The information may be needed to interpret changes in the data that occur as a result of forms or coding changes.

12.7 FLOW AND STORAGE OF COMPLETED DATA FORMS

Data forms should flow to data entry for keying and storage as they are completed. (See Chapters 16, 17, and 24 for additional discussion concerning data flow, editing, and storage procedures.) Continuous unrestricted flows are preferable to those that are constrained by batching requirements (e.g., such as those imposed by requiring a clinic to forward forms for processing only at specified time intervals).

Intermediate stops as a form moves from the clinic to the data center for processing should be avoided, if at all possible. Many of the Veterans Administration multicenter trials have procedures in which forms are sent from clinics to the

study chairman's office for a preliminary review and edit, and then to the data center for keying, editing, and storage. The intermediate stop delays receipt of the forms at the data center, thereby reducing the usefulness of the edits and analyses carried out by the center. Further, intermediate stops complicate communications with clinics concerning missed visits or deficient forms, since the inventorying and editing responsibilities are shared by the chairman's office and the data center.

The requirements for form storage should be addressed early in the course of the trial, ideally before any forms have been completed. The storage plan should be designed to protect the records from any unauthorized use and against loss or destruction. Protection of the latter type may require maintenance of duplicate files—one at the clinic and the other at the data entry site. Large or important files may be microfilmed to reduce the space required for storage or as a further safeguard against loss.

Part III. Execution

Chapters in This Part

The four chapters of this Part are concerned with execution of the trial. Chapter 13 outlines the steps required in executing the trial, with emphasis on the steps to be carried out in getting started. Chapters 14 and 15 concentrate on the recruitment, treatment, and follow-up processes. The last chapter details general procedures needed to ensure the quality of the data generated in a trial.

13. Preparatory steps in executing the study plan

The lame man who keeps the right road outstrips the runner who takes a wrong one. Nay, it is obvious that when a man runs the wrong way, the more active and swift he is the further he will go astray.

Sir Francis Bacon

13.1 ESSENTIAL APPROVALS AND CLEARANCES

All trials require completion of a series of steps before they can be started. The steps outlined in this chapter are in addition to those discussed in Chapters 11, 12, and 21 with regard to preparation of the study plan, data forms, and funding request.

13.1.1 IRB and other approvals[1]

One set of approvals has to do with those provided by the institutional review boards (IRBs) of individual centers in a trial (clinics, as well as the data center and any other resource center concerned with data collection or patient care). The main function of the board is to provide assurance that the proposed research meets accepted standards of ethics and medical practice. Technically, the assurance is needed only for federally funded studies. However, most institutions require reviews for all research involving humans, regardless of the source of funding. The impetus for the boards grew out of concerns in the 1960s regarding the nature and extent of research involving humans. A memo dated February 8, 1966, from the Surgeon General of the United States Public Health Service mandated creation of the local boards as a prerequisite for continued funding. The structure for IRBs, their composition, and their domain of responsibility has subsequently been spelled out in federal regulations on protection of human subjects (Office for Protection from Research Risks, 1983).

Each board, in order to comply with current regulations, must:

- Have at least five members
- Not be made up exclusively of members of one sex or of one profession
- Include at least one member whose primary concerns are in a nonscientific area (e.g., law, ethics, theology)
- Include at least one member who is not otherwise affiliated with the institution and who is not part of the immediate family of a person who is affiliated with the institution
- Exclude any member from review of a specific proposal who has a conflict of interest (e.g., is an investigator in a study under review)

Individual IRBs have their own rules regarding time schedules for submissions, formats for proposals, and the nature and amount of materials to be supplied. Table 13-1 lists the information requirements as envisioned for a "typical"

[1] See Section 14.6 for additional comments.

Table 13-1 Information required for IRB approval

- Statement of study objectives and rationale
- Description of the study treatments and methods of administration
- Recap of prior evidence concerning safety and efficacy of the study treatments
- Type and source of study patients
- Primary outcome measure for assessing the study treatments
- Length of patient follow-up
- Number of patients to be enrolled and rationale for proposed sample size
- Risk-benefit analysis of trial
- Method of treatment assignment (e.g., random, physician choice, etc.)
- Summary of methods for protecting patients from needless or prolonged exposure to a harmful study treatment
- Summary of safeguards to protect patient privacy and confidentiality
- Consent statement and related material

IRB in relation to clinical trials. Specifics will vary from board to board.

The material submitted to the IRB should indicate the nature and extent of safety monitoring to be performed (see Chapter 20). The individual or group responsible for this function should be identified in the submission along with sufficient details to enable members of the IRB to make an informed judgment regarding the statistical credentials and expertise of the individual or group named. The submission should include a general description of the methods to be used for safety monitoring, the frequency of interim analyses for monitoring purposes, and the procedure to be followed in communicating with local investigators and the IRB regarding proposed treatment changes emanating from the monitoring. Details regarding the communication process are especially important in trials in which monitoring responsibilities are vested in an individual or group that is not under the control of the local clinical investigator, as in most multicenter trials and some single-center trials.

The National Institutes of Health (NIH) will not review a research proposal involving humans without assurance from the proposing investigator's IRB. The assurance is supplied via completion of form HHS 596 (Protection of Human Subjects Assurance/Certification/Declaration) that is signed by a responsible official of the IRB.

Proposals for clinical trials may require at least two IRB reviews before initiation of patient intake. The first will be required in conjunction with the submission of the funding proposal to the sponsor. The second will be required after the proposal is funded and before the initiation of patient intake, after the details of the study protocol and consent process have been set.

The proposing investigator is responsible for communications with his IRB. He must be prepared to address their concerns in a forthright manner and to revise consent statements in accordance with their requests. Concerns regarding the rights of patients to privacy and confidentiality, as well as safety issues, must be addressed. The entire review and clearance process may take months and may be complicated by the need to clear changes through the leadership of the study, in the case of multicenter trials (see Section 14.6.2 for added details).

Additional reviews and approvals will be needed if the trial involves use of hazardous materials, such as radioactive isotopes, or laboratory animals.

13.1.2 IND and IDE submissions

Most drug trials will require submission of an Investigational New Drug Application (INDA, also referred to as an IND) to the Food and Drug Administration (FDA) before they can be started (Food and Drug Administration, 1981). Table 13-2, Part A, lists general items of information required for an INDA.

An INDA is required for any drug that is not approved by the FDA for the indication proposed. The requirement extends to established drugs that are to be used in ways that depart from prescribed practice, as indicated in the label insert. For example, the University Group Diabetes Program (UGDP) needed an INDA for both tolbutamide and phenformin even though they had been approved by the FDA as hypoglycemic agents. Even a nonprescription drug requires an INDA if it is used like a prescription drug. For example, one was required for aspirin in both the Coronary Drug Project Aspirin Study (CDPA) and Aspirin Myocardial Infarction Study (AMIS).

The FDA approval process can delay the start of the trial and lead to alterations in its design. Investigators in the National Cooperative Gallstone Study (NCGS) were required to carry out

Table 13-2 Items of information required for IND and IDE submissions to the FDA

A. Investigational New Drug Application (Summarized from FDA Form 1571, 10/82, Notice of Claimed Investigational Exemption for a New Drug)

- Details concerning the drug, including drug name, composition, source, method of preparation, quality control procedures in production and packaging
- Summary of previous investigations involving the drug
- Copies of informational material (including information on label and labeling) about the drug to be supplied to investigators involved in administering the drug
- Name and qualifications of each investigator to be involved in proposed studies
- Name and qualifications of personnel responsible for monitoring progress of proposed studies and for safety monitoring
- Description of the study plan, including details, in the case of proposed clinical trials, regarding sample size, duration of the study data collection, methods of treatment, as well as details concerning the IRB responsible for reviewing the proposed work, and details regarding informed consent
- Assurances from the IND sponsor that:
  - The FDA will be notified if the investigation is discontinued and of the reasons for the action
  - Each investigator associated with the IND will be notified if an NDA for the drug is approved, or if the investigation is discontinued
  - If the drug is to be sold, an explanation will be supplied to the FDA as to why sale is required and why sale should not be regarded as commercialization of the drug
  - Clinical studies in humans will not be initiated prior to 30 days after receipt of the Notice of Claimed Investigational Exemption for a New Drug by the FDA, unless otherwise indicated by the FDA
  - An environmental impact statement will be provided to the FDA, if so requested
  - All nonclinical laboratory studies have been or will be conducted in accordance with the Good Laboratory Practice regulations of the federal government, or that reasons why they have not or cannot be followed will be supplied to the FDA

B. Investigational Device Exemption (Summarized from reference 189, Appendix I)

- Name and address of sponsor of IDE along with names and addresses of all other investigators to be involved in the IDE
- Summary of prior investigations of the device
- Description of the methods, facilities, and controls used for the manufacture, processing, packaging, storage, and, where appropriate, installation of the device
- Certification that all investigators have signed an agreement to be involved in the IDE and that no new investigators will be added without signed agreements
- Name and address of the chairperson of each IRB associated with the IDE request
- Details regarding price of the device if it is to be sold and an explanation of why sale does not constitute commercialization of the product
- An environmental impact statement when requested
- Details concerning labeling of the device
- Copies of all forms and informational materials to be provided to patients in relation to the consent process
- Description of the study plan including:
  - Statement of purpose
  - Study protocol
  - Risk analysis
  - Description of the device
  - Methods for monitoring the investigation (progress as well as safety), including names and addresses of monitors

biopsy studies of patients treated with chenodeoxycholic acid before they were allowed to proceed with a full-scale trial of the drug (National Cooperative Gallstone Study Group 1981a, 1981b, 1984).

Amendments to the Federal Food, Drug, and Cosmetic Act of 1938, passed in 1976, extended the regulatory authority of the FDA to medical devices. A medical device is defined as (Food and Drug Administration, 1983):

> *Any instrument, apparatus, implement, machine, contrivance, implant, in vitro reagent, or other similar or related article, including component, part, or accessory, which:*
>
> - *Is recognized in the official National Formulary, or the United States Pharmacopeia, or any supplement to them;*
> - *Is intended for use in the diagnosis of disease or other conditions, or in the cure, mitigation, treatment, or prevention of disease, in man or other animals; or*
> - *Is intended to affect the structure or any function of the body of man or other animals; and*

- *Does not achieve any of its principal intended purposes through chemical action within or on the body of man or other animals and which is not dependent upon being metabolized for the achievement of any of its principal intended purposes.*

The definition covers approximately 1,700 devices that range from blood collection tubes and tongue depressors to heart valve replacement materials and pacemakers.

The FDA has established three classes of devices, based on the degree of control deemed necessary for assuring the safety and efficacy of the device (Food and Drug Administration, 1983). All three classes are subject to the Good Manufacturing Practices Regulations. In fact, the only controls required for Class I devices (e.g., capillary blood collection tubes, tongue depressors, crutches, and arm slings) are via these regulations. Added assurances for Class II devices (e.g., hearing aids, blood pumps, catheters, and hard contact lenses) and Class III devices (e.g., life-support or life-sustaining devices, such as pacemakers, intraocular lenses, and heart valve replacements, as well as devices considered of importance in preventing impairment of health) are provided via performance standards plus clinical trials for Class III devices. Permission to carry out trials of Class III devices is obtained via an Investigational Device Exemption (IDE), granted by the FDA. Part B of Table 13-2 lists items of information required in conjunction with an IDE application (Food and Drug Administration, 1980).

13.1.3 OMB clearance

The Office of Management and Budget (OMB), one of the offices in the executive branch of the United States government, has the authority to review and approve data forms used by all branches of the federal government, including the NIH. Technically, any data form to be administered or distributed to ten or more people that is produced by a governmental agency, or by a group under contract to it, requires OMB clearance—even draft versions of data forms developed simply for testing purposes. Forms developed under NIH grants are not subject to the order.

The review can delay the start of data collection, especially if staff at OMB regard certain forms or items as unnecessary or to constitute an invasion of a person's privacy. Usually, however, the review and approval process is not a major stumbling block. In fact, many areas of clinical investigations are exempt from review and those that are required may be achieved in short order if the project officer of the sponsoring agency maintains an effective working relationship with OMB staff and allows sufficient lead time for clearance.

13.2 APPROVAL MAINTENANCE

13.2.1 IRB

The approval granted by the IRB prior to the start of the trial and for each renewal will be for a one year period, unless otherwise indicated. The submission accompanying a renewal request should indicate the nature and extent of progress made since the initial request or last renewal request, the reasons for continuing the study, and proposed changes in the study protocol or consent procedures. Changes must be cleared before they can be implemented. Those that cannot wait for the annual review will require special reviews.

The IRB may require a synopsis of interim results for renewals of trials requiring safety monitoring (see Table 22-1 in Chapter 22). Complying with this request will pose problems in trials in which clinical investigators are denied access to interim results for reasons discussed in Chapter 22. The results portion of the renewal submission will have to be prepared and submitted by nonclinical personnel in such cases. The boards may be willing to forego looks at interim results if they are satisfied with the safety monitoring done in the study, as discussed in Section 13.1.1. They may have no choice in multicenter trials if clinics are not given access to interim results. Theoretically, they could still insist on synopses of results for the clinic in question, but they would be of little value because of the numbers of patients involved.

Investigators are obligated to report unexpected adverse events as they occur. Those reports are reviewed as they are received and may lead to immediate suspension or withdrawal of the approval until or unless changes mandated by the IRB are made.

13.2.2 FDA

The individual (or agency) to whom the INDA or IDE is granted is required to report unexpected adverse events to the FDA as they occur. There is also a requirement to provide summaries of study results as the trial progresses. The latter reporting requirement may be satisfied by simply supplying the FDA with copies of reports prepared for the treatment effects monitoring committee (see Chapter 23). Both the CDP and NCGS satisfied the majority of their FDA reporting requirements in this way.

13.2.3 Other approvals

Other approvals granted at the start of the trial, such as for use of radioactive compounds or controlled substances, will have to be updated as the trial proceeds. Changes to the data forms may have to be cleared through OMB if the study is funded via a government contract. Sponsoring agencies, such as the NIH, will require interim progress reports to continue funding for the trial.

13.3 DEVELOPING STUDY HANDBOOKS AND MANUALS OF OPERATIONS

Any trial requires two basic sets of documents: one that describes clinic operations and another that describes the data intake and processing procedures in the trial. These two sets of documents may constitute separate sections in the same handbook or manual or may be contained in separate documents (see Appendix G).

A large multicenter trial may require several other documents in addition to the two mentioned above. Studies with a central laboratory will need a document that describes its methods and procedures. Other resource centers, such as those needed for performing special reading or coding functions, will also need documents detailing their practices.

The groundwork needed for production of the required handbooks and manuals is laid when the trial is planned. The work involved in writing and maintaining these documents will start shortly after the trial is funded and continue until it is finished. Table 13-3 contains a list of suggestions concerning their development and maintenance.

A handbook, as used in this context, is a document that contains a series of tables, charts, figures, and specification pages that detail the design and operating features of the trial. A manual, as discussed herein, is a document that details the methods and procedures of the entire trial or some aspect of it largely through written narrative and accompanying tables, charts, and figures. The two kinds of documents serve somewhat different functions and, hence, are not necessarily interchangeable. The primary virtue of a handbook lies in its organization and in the tabular nature of the material presented. It is designed for use as a ready reference for study personnel. Manuals are designed to document procedures used in the trial. They are most useful to persons who want a detailed description of the actual procedures used.

The two kinds of documents may be developed simultaneously or in sequence, starting with the handbook. The latter approach was used in the Hypertension Prevention Trial (HPT). Work on the manual of operations was delayed until the handbook was developed. The development of the handbook simplified the task of preparing the study manual of operations. Further, the fact that the trial had been under way about 9 months when the work started allowed its developers to reference existing study documents and task specific manuals, thereby avoiding the need for inclusion of those details in the main document.

13.4 TESTING THE DATA COLLECTION PROCEDURES

Three general assurances should be satisfied before data collection is initiated:

- Essential data collection and patient examination procedures have been reviewed and approved by the study leadership
- Data forms needed for patient enrollment and for the initial phase of treatment and follow-up have been tested and are ready for use
- Projected time requirements for developing, testing, reviewing, and approving data collection procedures and related data forms for use in the later stages of treatment and follow-up are consistent with the data collection schedule of the trial

Satisfying the last condition may require a delay in the start of patient recruitment, even though the initial data intake procedures have been tested and approved. Once the first patient is enrolled, the rest of the data collection schedule is lockstep. It is better to delay the start of patient recruitment than to be forced into postponing follow-up visits because of the lack of

Table 13-3 Suggestions for development of study handbooks and manuals of operations

A. General

- Identify major topics or functions for which handbooks or manuals are required (e.g., clinic operations, data intake and processing, laboratory procedures, etc.)
- Develop a draft table of contents for each required handbook or manual and submit for review and comment by the leadership group of the trial before development
- Develop methods and procedures for data collection with input from key study personnel, including clinicians, statisticians, clinic coordinators, laboratory technicians, and the like
- Ensure that written material contained in handbooks or manuals is concise and devoid of complex sentences and esoteric language
- Test the adequacy of each handbook or manual by having it reviewed by individuals who will be using it
- Release a handbook or manual for use only after it has been reviewed and approved by the leadership of the study

B. Organization

- Each handbook or manual should have an official name and should be easily distinguished from all other handbooks or manuals in the study (e.g., through use of different colored binders)
- The name of the handbook or manual, date of release, version or edition number, and the name of the individual or group responsible for its distribution should be indicated on the title page
- Include a detailed table of contents, along with a listing of all tables and figures in the document
- Include a subject index and glossary
- Chapters in manuals should be divided into numbered subsections; the accompanying numbers and titles should appear in the table of contents of the document
- Left-hand page margins should be wide enough to keep text from being obscured or lost when pages are photocopied or bound (e.g., at least 1¼" for standard 8½ x 11" pages assembled in loose-leaf notebooks or pressure binders)
- Right-hand page margins should be wide enough to allow room for user notes (e.g., at least ¼" for standard 8½ x 11" pages). The same is true of top and bottom margins
- Pages should be typed using high resolution type fonts to allow for image degradation in photo-reproduction without a serious loss of legibility
- Boldface type, underlining, or other methods should be used to identify key phrases, definitions, and important procedural statements
- Ideally, pages should be numbered sequentially from the beginning to the end of a document, without regard to chapter or subsection. Numbering systems that recycle by chapter or section allow for page updates without disrupting the entire numbering system. However, such systems are not as convenient for users as are continuous numbering systems
- Placement of page and other identifying information should appear in a standard location on all pages (preferably upper right-hand corner) and should not be too near the edge of the page

C. Suggested maintenance aids

- Responsibility for periodic review and revision of a manual should be assigned to a specific individual or group
- A specific individual should be given responsibility for keeping track of revisions made to a handbook or manual and for making certain that all users of the handbook or manual are supplied with updates as they are produced
- Each new version of a handbook or manual should be identified with a revision date and should indicate the date and version number of the document it replaces
- Large documents that are subject to frequent updates should be kept in loose-leaf binders (facilitates page replacements and simplifies photo-reproduction of the document)
- Individual pages that are updated and inserted in an existing version of a document as replacements for outdated pages should include the revision date in the top or bottom right-hand corner of the pages

data forms or to use forms that have not been adequately tested.

The construction of the data collection instruments is one of the most important tasks in the entire study. General rules for item construction and forms development have already been discussed in Chapter 12. The paragraphs that follow deal with methods for testing the data forms.

It is probably fair to say that any item on a data form that can be misinterpreted will be. Some of the interpretation problems can be avoided by a careful review of all forms before any field testing is done. The next review should involve use of the forms on a few "practice" patients. Ideally, the forms should be completed for persons as similar to study patients as possible, but friends, colleagues, or spouses, instructed to behave and respond like "typical" patients, may suffice for some of the testing.

The entire set of data forms and accompanying procedures should be submitted to "walk-throughs," involving the staff who will be responsible for completing them, before they are

tested on real patients. The "walk-throughs" invariably identify items or sections that need to be relocated or rewritten to eliminate confusion or to streamline the way forms are to be completed.

Once these steps have been completed the forms are ready for field tests involving real patients. The test conditions should be as similar to those for the actual trial as feasible. The best approach is one in which the entire set of study procedures are carried out. However, this may not be possible for procedures that entail risks or that are justified only in special circumstances. The forms used should be in near final form, with one or two exceptions. They should make generous use of open-ended response categories, such as discussed in Section 12.5.4, in order to collect information useful in constructing response checklists for the final versions of the forms. They may also include alternative versions of the same item in order to determine the preferred wording of the item.

The number of patients used for the test should be large enough and heterogeneous enough to provide a reliable basis for preparation of final versions of the forms. The number will depend on available resources and on the complexity of the data collection scheme proposed. The penalties for undetected deficiencies are greatest in trials involving large numbers of patients.

The deposition of data collected in the test run should be settled before the run is undertaken if study-eligible patients are to be used in the test. The temptation in such cases is to reserve the option of adding the test data to the main data file if the number of changes mandated by the test is "small." The best approach is to preclude this option from the outset for several reasons. First, the desire to preserve the option may reduce the value of the test itself if investigators limit the changes they are willing to make simply as a means of maintaining the option. Second, the effort involved in merging test data into the main file may not be worth the return, especially if the merger requires a lot of recoding and reprogramming. Third, the absence of a stated policy can open the trial to criticism later on if the decision on use of test data appears to have been motivated by a desire on the part of the investigators to accentuate or ameliorate the observed treatment effect.

The intelligibility of any material that is read or given to patients in the trial should receive special scrutiny during the testing process. Particular attention should be paid to the patient consent statement and related materials. They should be tested on sample patients and then modified where necessary to ensure a clear and accurate presentation of the trial.

13.5 DEVELOPING AND TESTING THE DATA MANAGEMENT SYSTEM

Ideally, the development of computer programs needed to inventory completed data forms, and to edit, store, and retrieve data contained on those forms, should be started as soon as the forms have been developed for testing. However, this ideal is rarely achieved in reality. For one reason, even experienced investigators can underestimate the time required to develop a functioning data management system. Inexperienced investigators may not even recognize the need for one until well into the trial. Other reasons have to do with time and resource limitations. Of necessity, most of the work in the initial phase of a trial is devoted to development of the study protocol, data forms, and the data system. The pressures to complete these tasks and to get started with patient recruitment makes it difficult to find the time needed to develop a working data management system. The problem is compounded by the fact that it is not feasible to develop a working system until data collection procedures for the trial have been set—something that may not be done until patient recruitment is ready to start.

It is the responsibility of the data center to make sure that essential data management routines are available when needed. Basic routines, such as those needed for randomization, must be available by the time the first patient is randomized. Others, such as for inventorying data forms, should be available as soon as forms begin arriving at the center. The same is true for the editing routine to be applied to completed forms. Work on programs needed for performance and safety monitoring should begin soon thereafter (see Chapters 16 and 17).

The decision to start data intake before the data management system is in place can jeopardize its subsequent development. A good data center will keep this from happening by insisting on adequate lead time for its development before the start of data collection.

13.6 TRAINING AND CERTIFICATION

As a minimum, data collection personnel should be required to work through a sample set of data collection forms and to familiarize themselves with study procedures before being allowed to start data collection. Obviously, training cannot be started until data forms are in final form and needed documents, such as the study handbook or manual of operations, are available. This familiarization effort may be followed by workshops for demonstrating specific procedures and for observing personnel performing assigned data collection tasks. The training may be part of a formal certification process in which personnel are required to pass proficiency tests before they are allowed to start data collection, for example, as used in the DRS (Diabetic Retinopathy Study Research Group, 1981). This process should be started well before the projected start of data collection in order to avoid delays due to certification failures.

The training and certification processes are an essential part of quality control. They should be maintained over the course of data intake. Existing personnel should be required to undergo refresher training and recertification at intervals over the course of the trial. New personnel, recruited during the course of the trial, should be required to go through essential training and certification procedures before starting data collection in the trial.

The need for training and certification is most apparent in multicenter trials. Special efforts are required in such cases to make sure that all clinics are operating under the same ground rules and that they are adhering to established data collection procedures. However, the need is not unique to such trials. It extends to single-center trials as well. The opportunity for variation and misunderstanding with regard to data practices can be as great, sometimes greater, than in multicenter trials.

13.7 PHASED APPROACH TO DATA COLLECTION

Once the necessary testing and certification have been completed, patient enrollment may begin. There is a temptation, once this point is reached, to proceed as rapidly as possible. However, some initial restraint is wise, since live study conditions can be expected to reveal heretofore undetected defects. The larger the number of patients already enrolled when the defects are discovered, the greater the costs involved in correcting them.

A phased approach to data collection is especially important in multicenter trials involving a large number of clinics. Allowing all clinics to start data collection at the same time can swamp the data center before staff have had a chance to develop a functional data system. This problem can be minimized in one of two ways. One way is to fund only a skeleton set of clinics to begin with. The full complement of clinics can be recruited and funded once data collection is under way in the initial set of clinics. This approach was used in the CDP. The study started with just 5 clinical centers in 1965. A second set of 29 clinics was added in 1966. A third set of 21 clinics was added in 1967 to bring the total to 55 (Zukel, 1983).

The other way, when a full complement of clinics is identified from the outset, is to authorize only one or two clinics to start data collection. Other clinics are not phased in until essential support systems have been developed and tested. This approach was used in the Multiple Risk Factor Intervention Trial, MRFIT (Sherwin et al., 1981). The sponsoring agency must have the flexibility needed to determine when funding for data collection is to start in the individual clinics to make this approach viable.

14. Patient recruitment and enrollment

Seek, and ye shall find.

Matthew 7, verse 7

14.1 RECRUITMENT GOALS

The recruitment goal in fixed sample size designs should be set before the trial is started. As noted in Chapter 9, it may be based on a formal calculation or on practical considerations. It serves as a landmark for gauging progress during patient recruitment when accompanied by a timetable to indicate when it is to be achieved.

It is not uncommon for trials to fall short of their stated goal, even when the recruitment period is extended well beyond the date originally set for achieving the goal. The Coronary Artery Surgery Study (CASS) extended the recruitment time and even then enrolled fewer patients than originally planned. The same was true for the Program on the Surgical Control of Hyperlipidemia (POSCH). Their recruitment experiences are similar to those outlined for trials carried out as part of the Veterans Cooperative Studies Program (Collins et al., 1980). Unfortunately, it is not easy to assess the recruitment performance of many of the completed trials because of the absence of details in published reports concerning the original recruitment goal and timetable for achieving it.

Investigators may set a number of secondary recruitment goals or quotas in addition to the main one. Some may relate to the mix of patients within a clinic (e.g., the number of males versus females). Others, in the case of the multicenter trials, will relate to the numbers of patients to be enrolled per clinic. All secondary goals should be viewed as general guidelines rather than as absolute for practical reasons. For example, it is more efficient to allow all clinics in a multicenter trial to recruit to a common cutoff date than to a set number per clinic. The same is true with regard to goals or quotas regarding the mix of patients within a clinic. Certain kinds of patients will be harder to find than others. Insistence on a specified mix will increase the time needed for patient recruitment.

14.2 METHODS OF PATIENT RECRUITMENT

Table 14-1 lists methods of patient recruitment. The methods have been divided into those that rely on direct patient contact and those that do not. Each method has specific strengths and weaknesses that must be considered when a choice is made among them (Table 14-2). Any method of recruitment requires the support of colleagues to succeed. An investigator should not undertake a trial without this support.

Studies relying on patient referrals can expect to experience difficulties meeting their recruitment goal if referring physicians are not in sympathy with the study or if they are reluctant to make referrals for fear of "losing" their patients to the study. The National Eye Institute distributed letters to ophthalmologists announcing the start of the Diabetic Retinopathy Study (DRS),

Table 14-1 Methods of patient recruitment

Recruitment method	*Trials using method**
A. Direct patient contact	
• Clinic contacts	AMIS, CDP, UGDP
• Screenings	HDFP, MRFIT
• Direct mailings	HPT, LRC
B. Indirect patient contact	
• Referring physicians	AMIS, CASS, CDP, DRS, MPS, UGDP
• Retrospective record reviews	POSCH, UGDP
• Spot radio and TV ads	AMIS, MRFIT

*See Glossary for name corresponding to acronym.

Table 14-2 Comments concerning the choice of recruitment methods

Recruitment method	*Comments*
A. Direct patient contact	
Via primary care clinic	• Clinic must be large enough to yield the required number of patients if it is to serve as sole source of patients • The study investigator should be responsible for the primary care clinic or play a major role in its operation • Fellow colleagues in the clinic must subscribe to the tenets of the study and be willing to follow the prescribed treatment • Generally, only viable for relatively common diseases or conditions. Not viable if most patients seen at the clinic are ineligible for the study
Via screening	• Method of choice for identification of patients with a disease or condition that can be diagnosed with a simple and inexpensive test and that is not routinely diagnosed via regular patient care channels • May be used to supplement other recruitment methods when the disease or condition of interest is rare (e.g., a certain type of hyperlipemia) • Study clinic should have facilities to treat identified patients or must be prepared to refer patients not suitable for study to appropriate sources for care
Via mailings or telephone calls	• Best limited to recruitment for primary prevention trials or trials focusing on treatment of a disease or condition not presently being treated by the medical community • Not recommended for recruitment of patients with a disease or condition routinely diagnosed and treated. Direct appeals in this case may be viewed as efforts to "steal" patients • Method usually used in combination with screening procedures carried out at the clinic to determine the eligibility of those who respond to the direct mail or phone appeal. Screening is essential if a respondent is not likely to know whether he has the disease or condition of interest

Table 14-2 Comments concerning the choice of recruitment methods (*continued*)

Recruitment method	*Comments*
B. Indirect patient contact	
Via referring physician	• Required mode of recruitment if study clinic located in tertiary care facility. May be used as the primary method of recruitment or as an adjunct to other methods • Study clinic should be located in an established referral center for the disease or condition of interest • Patient's primary care must be compatible with study tenets • Not a reliable method of recruitment if the disease or condition is routinely treated by a primary care physician • Method works best for a disease or condition for which there is no recognized form of therapy and when the referring physician has no concern about "losing" referred patients • It may be necessary to augment the referral process by: – Mailing letters to referring physicians to inform them of the study and of the type of patients needed – Journal articles outlining the design and purpose of the trial – News articles in the medical or lay press concerning the trial – Presentations at medical meetings to acquaint referral physician with the trial
Via retrospective record reviews	• May be preferred method for rare disease or condition, if routinely diagnosed and noted in clinic records • Not useful if newly diagnosed patients are required, or where most patients identified by the reviews are likely to be ineligible for enrollment (e.g., because they have received a form of treatment that disqualifies them from consideration) • May have to be used: – When it is impractical or too costly to mount a screening effort to identify patients – When there is no risk-free low-cost screening procedure available – If eligible patients are unlikely to be referred to the study clinic – If the disease or condition is so rare as to make it impractical to consider any of the recruitment methods outlined above
Via radio or TV spot ads and the news media	• Usually used as an adjunct to other methods of recruitment • Often used to acquaint members of the lay and medical community with the trial

which outlined the type of patients desired for the trial. Care was taken in the letter to note that patients who were referred for study would remain under the care of the referring ophthalmologist for their regular eye care.

Some trials have used the news media to facilitate patient recruitment. Recruitment publicity may take the form of news stories appearing in area newspapers, may be aired on radio or television, or may consist of paid advertisements

MP-100
14146

aimed at certain types of patients. Some of the clinics in the Multiple Risk Factor Intervention Trial (MRFIT) used spot television ads to inform potential study candidates of the trial. Such direct appeals are only practical in settings where patients can be expected to know they have the disease or condition of interest and are not under treatment for it (see Chapter 24 for further discussion of study information policy issues). The need to have newly diagnosed, untreated patients can be a major stumbling block to recruitment if most of the patients arriving at a clinic are already under treatment. This was one of the difficulties in recruiting patients in the University Group Diabetes Program (UGDP).

Studies may be forced to establish their own screening and referral procedures if existing sources of patients are inadequate. Various trials, such as the Hypertension Detection and Follow-Up Program (HDFP), Coronary Primary Prevention Trial (CPPT) of the Lipid Research Clinics (LRC), and MRFIT, had to develop special screening procedures to find suitable patients. The LRC had to make over 436,000 patient contacts in order to find the 3,810 ultimately enrolled into the CPPT (Lipid Research Clinics Program, 1982). The MRFIT screened over 361,000 to find the 12,866 enrolled in that study (Multiple Risk Factor Intervention Trial Research Group, 1982). The HDFP screened over 158,900 to identify the 10,940 patients enrolled in that trial (Hypertension Detection and Follow-Up Program Cooperative Group, 1979a).

The systematic review of hospital records can offer a useful means of patient identification if the records can be expected to contain the needed information. However, it is not useful if most of the patients are ineligible because of their disease history or treatments received. The review is fairly easy to carry out if it is restricted to the investigator's own institution, but not if it involves other institutions as well, as in POSCH. That study relied on record searches at several hundred different hospitals. Special personnel were required to negotiate the agreements needed to make the searches (Matts et al., 1980).

14.3 TROUBLESHOOTING

The period of patient intake is crucial in the life of a trial. Special efforts are needed over the entire period to spot and correct problems that impede patient intake. Recruitment performance should be monitored closely by comparing the rate of enrollment with that required to achieve the stated recruitment goal in the time period specified. An extremely low recruitment rate may call for a relaxation of some of the selection criteria or cancellation of the entire study or of support for one or more of the clinics in it. The monitoring process may be facilitated by screening logs. The logs may help to pinpoint reasons for exclusions and, hence, may suggest ways of modifying selection criteria to increase patient yield. They may also help to characterize the ways in which the population enrolled differs from the population screened, as in CASS information that may be useful when generalizing results of the trial (Coronary Artery Surgery Study Research Group, 1984; see also Question 9 in Chapter 19).

Study leaders should conduct formal visits to clinics for on-site inspections. The first round of visits should be as soon after the start of patient recruitment as possible. Subsequent visits may be carried out at intervals over the life of the trial (see Section 16.8.3). The visits can be helpful in identifying and correcting problems and in bolstering the morale of clinic staff (see Cassel and Ferris, 1984, for discussion of site visiting procedures in the Early Treatment Diabetes Retinopathy Study, ETDRS).

14.4 THE PATIENT SHAKE-DOWN PROCESS

The process of evaluating a patient for entry into a trial may require several examinations. The longer the evaluation period, the easier it will be to identify uncooperative or otherwise unsuitable patients. Patients who fail to keep appointments or who do not comply with data collection requirements for baseline visits are not likely to become more compliant after enrollment.

Some drug trials (e.g., the CDP, Coronary Drug Project Research Group, 1973a) require use of a single-masked placebo during the prerandomization evaluation period to help identify noncompliant patients (see Question 37, Chapter 19). No medication, not even a placebo, should be given without explanation. Of necessity, the explanation must be less than forthright if clinic staff are to conceal its nature in the case of single-masked placebos. The evasive nature of the explanation required can strain the patient-physician relationship at a crucial point in the enrollment process.

14.5 THE ETHICS OF RECRUITMENT

The methods used for recruitment should be devoid of any procedures that may be construed as coercive. Cash payments as inducements for enrollment or for patients to continue in a trial should be used with caution, especially if the trial involves risks. They may be necessary in trials involving healthy volunteers who will not realize any direct benefit from the trials, but not in trials involving treatment of some health condition. In those cases, the benefits derived from the care provided should serve as a sufficient inducement for enrollment.

The recruitment process should not involve any restrictions on the demographic, social, or ethnic characteristics of the patient population, except those needed for scientific reasons (e.g., restriction of age to allow concentration on a high-risk group of patients, or restriction to the sex group with the preponderance of the disease) or for practical or ethical reasons (e.g., exclusion of non-English-speaking patients because of concern regarding adequacy of the informed consent process). However, this is not to say that the study may not end up with a preponderance of one sex or ethnic group, or with patients largely from the same social class. The composition will depend on patient sources available to clinics.

The recruitment procedures used in a trial may come under scrutiny long after enrollment has been completed. The Tuskegee Syphilis Study is a case in point (Schuman et al., 1955; Tuskegee Syphilis Study Ad Hoc Advisory Panel, 1973; Vonderlehr et al., 1936). Critics of the study have suggested that the concentration on poor, uneducated blacks led to a climate of complacency in the way it was run (Brandt, 1978; Jones, 1981; Rothman, 1982).

14.6 PATIENT CONSENT[1]

14.6.1 General guidelines

It is unethical to carry out any experiment that entails risks to humans without their voluntary consent. The Nuremberg Code[2] and all codes since then have been explicit on the need for voluntary consent (Levine and Lebacqz, 1979; Levine, 1981). However, relatively little attention was devoted to the actual consent process in medical research until the Surgeon General of the United States Public Health Service (USPHS) addressed the issue in a memo (dated February 8, 1966) to heads of institutions conducting research under Public Health Service grants. The memo ultimately led to detailed regulations, including the creation of institutional review boards (IRBs), as a means of ensuring adherence to ethical practices in the design and conduct of research on humans. Table 14-3 provides a summary of the pertinent points concerning the consent process, as contained in the most recent set of regulations. The regulations read in part:

> *Except as provided elsewhere in this or other subparts, no investigator may involve a human being as a subject in research covered by these regulations unless the investigator has obtained the legally effective informed consent of the subject or the subject's legally authorized representative. An investigator shall seek such consent only under circumstances that provide the prospective subject or the representative sufficient opportunity to consider whether or not to participate and that minimize the possibility of coercion or undue influence. The information that is given to the subject or the representative shall be in language understandable to the subject or the representative. No informed consent, whether oral or written, may include any exculpatory language through which the subject or the representative is made to waive or appear to waive any of the subject's legal rights, or releases or appears to release the investigator, the sponsor, the institution or its agents from liability for negligence (Office for Protection from Research Risks, p. 9, 1983).*

The requirement for consent, when first introduced, led to fear that it would make recruitment of patients for studies impossible. This fear has not been justified, although the burden imposed by the regulations is unfair in one regard. An investigator is required to make certain that a patient about to enter a trial understands the nature of the risks and benefits that may accrue from the treatments to be offered. Yet that same patient, when seen by his regular physician, may be offered similar treatments without any discussion of their risks or benefits (Chalmers, 1982a).

1 See Section 13.1.1 for additional comments.

2 The code was an outgrowth of the war crimes trials in Nuremberg following World War II. The code is reproduced in Levine, 1981.

Table 14-3 General elements of an informed consent

- A statement that the study involves research, an explanation of the research and the expected duration of the subject's participation, a description of the procedures to be followed, and identification of any procedures that are experimental
- A description of any reasonably foreseeable risks or discomforts to the subject
- A description of any benefits to the subject or to others that may reasonably be expected from the research
- A disclosure of appropriate alternative procedures or courses of treatment, if any, that might be advantageous to the subject
- A statement concerning the extent, if any, to which confidentiality of records identifying the subject will be maintained
- For research involving more than minimal risk, an explanation as to whether any compensation or medical treatments are available if injury occurs and, if so, what they consist of, or where further information may be obtained
- An explanation of whom to contact for answers to pertinent questions about the research and research subjects' rights, and whom to contact in the event of research-related injury to the subject
- A statement that participation is voluntary, refusal to participate will involve no penalty or loss of benefits to which the subject is otherwise entitled, and the subject may discontinue participation at any time without penalty or loss of benefits to which the subject is otherwise entitled
- When appropriate, one or more of the following elements of information shall also be provided to each subject:
  - A statement that the particular treatment or procedure may involve risks to the subject (or to the embryo or fetus, if the subject is or may become pregnant) that are currently unforeseeable
  - Anticipated circumstances under which the subject's participation may be terminated by the investigator without regard to the subject's consent
  - Any additional costs to the subject that may result from participation in the research
  - The consequences of a subject's decision to withdraw from the research and procedures for orderly termination of participation by the subject
  - A statement that significant new findings developed during the course of the research that may relate to the subject's willingness to continue participation will be provided to the subject
  - The approximate number of subjects involved in the study

Source: Reference citation 365.

14.6.2 The consent process

Table 14-4 provides a list (prepared by the author) of items that should be covered in the consent process. It differs from the list in Table 14-3 in that it is specific to the area of clinical trials. Appendix E contains sample consent statements from three of the trials sketched in Appendix B.

The consent process, to be valid, must be based on factual information presented in an intelligible fashion and in a setting in which the patient, or his guardian, is able to make a free choice, without fear of reprisal or prejudicial treatment. Meeting these conditons may be impossible in cases where the patient is highly vulnerable, either because of his medical condition or physical surroundings. Extra precautions are needed whenever minors, mental patients, or prisoners are approached. The class action suit for damages brought against investigators at the University of Maryland on behalf of Maryland state prisoners had to deal with questions concerning the nature of free consents obtained in prison settings (United States District Court for the District of Maryland, 1979). No damages were awarded, but the suit took years to complete.

Reservations concerning the adequacy of the consent process in institutionalized populations have all but eliminated these populations as patient sources for research studies. They have also tended to discourage trials in children. The latter trend is unfortunate. Some trials must be done in children to obtain information pertinent to their illnesses or treatments.

The consent process must be completed before the treatment assignment is issued (except with the method proposed by Zelen; see Section 14.8). No patient should be randomized who expresses a reluctance or unwillingness to accept whatever treatment is assigned. The process should include an explicit statement regarding a patient's right to withdraw from the trial at any time after randomization. The statement may be balanced with a discussion of the effect withdrawals have on the trial and the responsibility a patient has, within limits, to continue in the trial if he decides to enroll (Levine and Lebacqz, 1979).

It is best to avoid exact time specifications regarding the anticipated length of follow-up in long-term trials. The time, even if seemingly fixed at the outset, may have to be extended later for reasons unanticipated at the outset. Similarly, promises as to when the study treatment will be offered to patients assigned to the control treatment should be avoided if there is any chance of having to renege on them later, as was the case in the NCGS (National Cooperative Gallstone Study Group, 1981a).

Table 14-4 Suggested items of information to be imparted in consents for clinical trials

General descriptive and design information

- Description of the disease or condition being studied and how the patient qualifies for the study
- Type of patients being studied and the number to be enrolled
- Anticipated length of follow-up
- Description of data collection schedule and procedures

Treatment information

- List of the treatments to be studied and rationale for their choice
- Treatment alternatives available outside the study
- Nature of the control treatment
- Method of treatment administration
- Method of assigning patients to treatment
- Level of treatment masking
- Nature of information regarding treatment results that will be made available to patients during and at the conclusion of the trial

Risk-benefit information

- Description of the risks and benefits that may accrue to a patient from participation in the trial
- Enumeration of the potential risks and benefits associated with the study treatments, as well as an enumeration of common side effects
- Description of any special procedures that will be performed, including an enumeration of the risks and benefits associated with those procedures, and the time points at which they are to be performed

Patient responsibilities and safeguards

- Outline of responsibilities of patients enrolled in the trial, including discussion of the importance of continued follow-up
- Outline of what is expected of the patient in following the examination schedule and in carrying out special procedures between visits
- Outline of safeguards to prevent continued exposure of a patient to a harmful study treatment or denial of a beneficial one
- Outline of safeguards for protecting a patient's right to privacy and confidentiality of information
- Indication of a patient's right to withdraw from the trial at any time after enrollment without penalty or loss of benefits to which he is otherwise entitled
- Statement of the policy of the investigator's institution on compensation for, or treatment of, study-related injuries
- Statement of the patient's right to have questions answered regarding the trial and indication of items of information that will not be disclosed (e.g., the treatment assignment in a double-masked trial)
- Statement of the length of time personal identifiers will be retained after the close of the trial, where such information will be retained, and the reasons for keeping it (e.g., for use in contacting or recalling the patient after close of the trial). Statement should also indicate ways in which the information may be used (e.g., to access the National Death Index or other information sources for determining mortality status after the close of the trial)

Most clinical trials involve the collection and storage of personal information, such as name and address, on study patients (see Section 15.3 for uses of the information in tracing patients). Some investigators engaged in epidemiological studies have indicated the exact date at which such information will be purged from patient files. The commitment is unwise in long-term clinical trials for two reasons. First, it may be impossible to meet because of unexpected delays in the conduct of the trial. Second, and more important, there may be a need to contact patients after the trial is completed, especially if any of the study treatments appear to be producing late and unexpected adverse effects.

The mechanics of obtaining the informed consent must be individualized to the population to be studied. Information may be presented in various ways so long as there is adequate opportunity for a patient (or his guardian) to have all questions regarding the study answered before he is asked to make a decision on enrollment. Hard sells are to be avoided. First, because they represent subtle forms of coercion. Second, because they can lead to enrollment of uncooperative patients.

Whenever feasible, it is wise to carry out the consent process in two stages with a time separation of a day or more between the first and second stages. Many trials lend themselves to this approach, especially those that require multiple visits to establish a patient's eligibility for enrollment. Exceptions are cases in which treatment must be started on the spot.

The first stage should be designed to acquaint the patient with the study and its requirements. It should involve a conversation with the patient in a setting that is conducive to a two-day exchange. The information imparted should be supplemented with written material, including a copy of the consent statement for the patient to take home to review at his leisure. The second stage should be used to answer questions raised by the patient and to review what would be required of him if he agrees to enroll. The consent statement should be signed at the end of this stage.

Both stages should allow ample opportunity for the patient to question clinic personnel regarding the study and his role in it. A patient should not be asked to sign the consent statement if he has any doubts about enrolling or if the clinic staff believes he does not understand what his participation would involve. The patient should be asked to reaffirm his willingness to accept whatever treatment is assigned before he signs the statement.

The time point at which the consent process is initiated is important. If it is initiated too early in the recruitment process, a good deal of time may be wasted explaining the trial to individuals who are subsequently found to be ineligible for enrollment on medical grounds. However, delaying the start of the process until the eligibility assessment is complete may not allow enough time for an orderly two-stage consent, especially if there is any urgency to start treatment once eligibility has been established.

The treatment assignment should be issued on the same day the consent is signed. The treatment should be initiated as soon thereafter as feasible, preferably on the day of assignment. A large time gap between consent and initiation of treatment will tend to increase the patient's anxiety regarding treatment and may increase the chance of his withdrawing before treatment is started.

The consent statement used in multicenter trials should be standardized to the extent possible. Some variation in language may be unavoidable because of local IRB wording requirements. However, the amount can be minimized by providing clinics with a prototype statement that covers the items listed in Table 14–4. Individual clinics may not reduce or abridge information contained in the statement, but may add to it if required to do so by local IRBs.

14.6.3 Documentation of the consent

Federal regulations require that:

> *Informed consent shall be documented by use of a written consent form approved by the IRB and signed by the subject or the subject's legally authorized representative. A copy shall be given to the person signing the form* (*Office for the Protection from Research Risks, p. 10, 1983*).

The IRB must approve the consent statement and will want to review all information (written as well as verbal) presented to patients in conjunction with the consent process. The statement presented for signature may contain a written description of all pertinent information needed in the consent process, or may refer to materials presented orally or in an accompanying document, such as in a patient information booklet. The patient's signature should be witnessed by a third party, regardless of how the presentation is made. The patient should be given a copy of the consent form after it has been signed. The original should be kept in the patient's file.

The responsibility for obtaining informed consent goes beyond the simple mechanics of presenting and signing documents. It is the responsibility of all those connected with the study to ensure that the process is carried out in a responsible manner. This responsibility extends beyond the clinics in multicenter trials. The approved statements should be collected by the coordinating center for review and storage. The review should be done by the study leadership and should be aimed at making certain that the statements meet study standards. In addition, the center should set up procedures to withhold treatment assignments until signed consents have been obtained.

Clinic site visits (see Section 16.8.3) should include checks on the consent process. This can be done via a walk-through for a hypothetical patient or by witnessing the process being carried out with an actual patient. The visiting team may also talk to patients who have gone through the process to learn what they know about the trial. The Beta Blocker Heart Attack Trial (BHAT) assessed the quality of the consent process by interviewing a sample of patients (Howard et al., 1981).

14.6.4 What constitutes an informed consent?

The question of what constitutes an informed consent is complex. It depends on the information to be conveyed and on how it is perceived by the patient. The formal nature of the doctor-patient relationship, coupled with the patient's anxieties regarding his condition, can be major blocks to meaningful communication. Studies of the consent process suggest that patients may fail to comprehend much of what they are told (Howard et al., 1981).

Consent materials should be simply written. It is important for design concepts, such as randomization, placebos, and masking, to be explained in lay terms. Some investigators have chosen to exclude patients who do not comprehend fundamental aspects of the study design. The Hypertension Prevention Trial (HPT) required patients to correctly answer a series of questions on the trial before they could be enrolled. A vaccine study research group at the University of Maryland requires volunteers to pass a test on the trial prior to enrollment (Levine, 1976; Woodward, 1979).

The failure to cover important items of information in the consent statement can cause a dilemma later on. A case in point is the failure to specify the nature of follow-up that will be carried out on indivduals who drop out after enrollment. It is common in a long-term trial to employ special procedures to obtain up-to-date mortality data on all study patients, including dropouts, at the time of final data analysis (see Chapter 15). Normally, these procedures are carried out unobtrusively. Nevertheless, the preferred approach is to make the patient aware of the ways in which his personal identifying information may be used for tracing and mortality follow-up before he is enrolled. A patient who is uncomfortable with what is proposed should not be enrolled.

14.6.5 Maintenance of consents

Consents given at the time of enrollment may have to be updated to remain valid. Patients should be informed of any decision or action that is likely to affect their willingness to continue in the trial, such as a decision to stop a study treatment in another group of patients because of an adverse effect or to add a data collection procedure that is inconvenient, uncomfortable, or risky. The CDP informed all study patients of the decision to terminate use of the high-dose estrogen treatment, even though less than one-quarter of them were on that treatment.

Changes in the federal regulations regarding the informed consent process during a trial may require addendums to the consents. For example, investigators in the UGDP were required to obtain signed consent statements from patients after recruitment had been completed. Undocumented oral consents, obtained at the start of the trial, were not considered sufficient after the February 8, 1966, memo from the Surgeon General of the USPHS. More recently, addendums have been required to inform patients of local policy on compensation for and care of study related injuries.

14.7 RANDOMIZATION AND INITIATION OF TREATMENT

Patients judged eligible and who are willing to participate in the trial are ready for enrollment. The point at which the treatment assignment is disclosed to the treating physician should be used to mark formal entry of a patient into the trial. Once enrolled, a patient should be counted as part of the study population (see Chapter 18).

The randomization procedure should be set up to make certain that assignments remain masked until they are needed for initiation of treatmentt (see Chapters 8 and 10). As already noted in Section 14.6.2, treatment should be started as soon after enrollment as practical, ideally on the day of randomization.

14.8 ZELEN CONSENT PROCEDURE

The usual approach is to obtain a patient's consent before he is randomized. The sequence is reversed in a modification proposed by Zelen (1979). In that method, eligible patients are randomized before consent is obtained. Those assigned to the control (standard) treatment are given that treatment without discussion of the alternative treatment(s) under evaluation. Only patients assigned to the test treatment(s) are given an opportunity to refuse the treatment assignment. Patients who refuse are given the control treatment.

The appeal of the approach lies in the fact that only patients assigned to test treatments are presented with information on treatment alternatives. The others are spared the anxiety that may be aroused by such discussions. However, in actual fact, most IRBs are reluctant to accept the approach, except under very special circumstances (such as in a trial involving a high-risk treatment on patients with a poor prognosis for life), and then only where cogent arguments can be made in its favor.

The approach has a number of limitations. Of necessity, it is limited to unmasked trials since the treating physician must know the assignment to identify patients with whom choices are to be discussed. In addition, refusals after randomization, if sizable, will make it difficult to reach any conclusion from the trial. Further, the procedure can lead to subtle forms of coercion. Patients assigned to the test treatment may be coaxed by study personnel to accept the assignment simply as a means of avoiding the data analysis and interpretation problems that can arise if there are a lot of treatment refusals. Finally, the method is unfair in that only patients assigned to test treatments are allowed a choice.

15. Patient follow-up, close-out, and post-trial follow-up

There are only two classes of mankind in the world—doctors and patients. . . . you [doctors] have been, and always will be exposed to the contempt of the gifted amateur—the gentleman who knows by intuition everything that it has taken you years to learn.

Rudyard Kipling

15.1 INTRODUCTION

Before approaching the subject matter of this chapter it is necessary to provide working definitions of three different processes. They are:

Patient follow-up
A process involving periodic contact with the patient after enrollment into the trial for the purpose of administering the assigned treatment, observing the effects of treatment, modifying the course of treatment, and collecting data to evaluate the treatment.

Patient close-out
A process carried out to separate a patient from the trial, involving cessation of treatment and termination of regular follow-up.

Patient post-trial follow-up
A process that involves patient follow-up after completion of the close-out stage of the trial and that is designed to yield information on the primary or a secondary outcome measure.

This chapter deals with the steps involved in carrying out these three processes (see also Appendix D).

15.2 MAINTENANCE OF INVESTIGATOR AND PATIENT INTEREST DURING FOLLOW-UP

The follow-up process requires a dedicated and committed staff to schedule and carry out the required examinations and a willing patient population. Both are needed if the trial is to succeed.

15.2.1 Investigator interest

Investigator commitment to the trial and interest in its activities must be high throughout if it is to succeed. Interest will be easy to maintain in a short-term trial, where the initial enthusiasm that usually accompanies the start of any new activity is enough to carry it through to completion. However, even in such cases spirits can sag before the data analyses are done and the final paper has been written. They can sag long before that point in long-term trials. Table 15-1 lists some of the aids that can be used to maintain investigator interest. The list is written with long-term multicenter trials in mind. However, morale problems are not unique to multicenter trials. They can be just as great in single-center trials.

Periodic meetings of study personnel are essential in maintaining a cohesive investigative group. They are needed before the start of the trial to outline the treatment and data collection procedures for the trial, and they are an essential

Table 15-1 Aids for maintaining investigator interest

- Periodic meetings of all study personnel
- Distribution of periodic progress reports on patient recruitment and follow-up, data collection, and other performance characteristics of the trial for review by all members of the investigative group
- Periodic newsletters distributed to study personnel designed to inform them of study progress, protocol changes, and so forth
- Investigator participation in the analysis of results and in writing or presenting papers concerning the trial
- Preparation of reports and papers during the course of the trial summarizing the design, organizational, and operating features of the trial
- Execution of ancillary studies
- Certificates of appreciation from the sponsor, and signed by key study leaders, to staff reaching important milestones (e.g., their five-year anniversary with the study)

part of the quality assurance process once it is under way. Meetings should include clinic coordinators, technicians, and other support staff important to the trial, as well as senior personnel.

The long-term multicenter trial will require a variety of other ways to maintain investigator interest. The chance for investigators to engage in ancillary studies (see Glossary for definition and Section 22.7.3 for discussion of management issues related to such studies) can help maintain their interest and general commitment to the trial. The opportunity to carry out analyses on data collected during the trial can also help morale. In reality, the opportunities for such analyses may be limited in settings in which there is a desire to mask clinic staff to treatment results, as discussed in Chapter 22. However, this policy does not preclude access to data unrelated to treatment outcome. The Coronary Drug Project (CDP) allowed access to baseline data for all the treatment groups as well as follow-up data for the placebo-treated group of patients. The follow-up data were used to generate several papers on the natural history of coronary heart disease (see Table B-3 of Appendix B for list).

Access to adherence or process measures by treatment group is also acceptable. Staff in the Multiple Risk Factor Intervention Trial (MRFIT) were provided with data indicating the level of risk reduction achieved as the study progressed. These summaries included data on clinic performance in terms of achieving stated treatment goals and were used by study leaders to assess the intervention procedures.

15.2.2 Patient interest

A patient's interest in the trial and willingness to continue in it can be expected to diminish with time. The longer the period of follow-up the greater the need for measures to counteract waning interest and participation levels. Table 15-2 lists factors and approaches that can help sustain patient interest in the trial. However, by all odds the most important factor is the attitude of clinic staff. Uninterested or discourteous staff will lead to an uninterested patient population.

15.3 LOSSES TO FOLLOW-UP

A loss to follow-up occurs whenever an item of information required as part of a scheduled follow-up examination is not obtained in the permissible time window (see Glossary for definition). The loss may be due to:

- Failure of the clinic staff to complete an item on an otherwise properly completed data form
- Failure of the patient to agree to certain procedures during an examination
- Failure of the patient to return to the clinic for an examination within the time window specified for it

Losses due to missed examinations or to examinations that are not done within the specified time window, and hence are counted as missed, are more worrisome than the losses resulting from failure to complete specific items or procedures during an examination. Further, an occasional missed examination for a patient has different implications than does a sequence of missed examinations. The longer the sequence, the greater the uncertainty regarding the outcome status of the patient.

Patients who are no longer able or willing to return to the clinic for scheduled follow-up examinations are dropouts. The declaration may be made by the patient (e.g., by announcing an intent to leave the study because of a lack of interest or because of a forthcoming move to another city) or by clinic staff. The latter will be the case with a patient who disappears or who does not, for whatever reason, keep his scheduled appointments. However, a clinic declaration should not be made until (1) clinic staff have made a concerted effort to locate the patient if

Table 15-2 Factors and approaches that enhance patient interest and participation

- Clinic staff who treat patients with courtesy and dignity and who take an interest in meeting their needs
- Clinic located in pleasant physical surroundings and in a secure environment
- Convenient access to parking for patients who drive, and to other modes of transportation for those who do not
- Payment of parking and travel fees incurred by study patients
- Payment of clinic registration fees and costs for procedures required in the trial
- Special clinics in which patients are able to avoid the confusion and turmoil of a regular out-patient clinic
- Scheduled appointments designed to minimize waiting time
- Clinic hours designed for patient convenience
- Written or telephone contacts between clinic visits
- Remembering patients on special occasions, such as Christmas, birthday anniversaries, etc.
- Establishment of identity with the study through proper indoctrination and explanation of study procedures during the enrollment process; through procedures such as use of special ID cards to identify the patient as a participant in the study, and by awarding certificates to recognize their contribution to the trial

he has disappeared, and to try to convince him to return to the clinic for a follow-up examination; and (2) the patient has missed a specified number of follow-up clinic visits. The date of the patient's last completed follow-up examination should be used as the date of dropout. The patient should remain classified as a dropout until or unless he returns to the clinic for a follow-up examination.

Patients who are classified as dropouts may or may not be lost to follow-up for the outcome of interest. They are when the diagnosis or measurement of the primary outcome can only be done at follow-up examinations performed in study clinics. They are not when it can be done outside study clinics (e.g., as in trials with death as the primary outcome). Similarly, conversion of a patient from active to dropout status may or may not affect his treatment compliance (see Glossary for definition). It will not if the conversion occurs after treatment has been completed and if the treatment cannot be reversed or nullified. It will be tantamount to creating a state of noncompliance if the conversion requires termination of an ongoing treatment process (e.g., as in most chronic drug treatment trials).

The willingness of a patient to remain under active follow-up will depend on a variety of factors, including:

- The amount of time and inconvenience involved in making follow-up visits to the clinic
- The perceived importance of the procedures performed at follow-up visits from a health maintenance point of view
- The potential health benefits associated with treatment versus potential risks
- The amount of trauma and discomfort produced by the study treatment or procedures performed
- The number and type of side effects associated with treatment

The dropout rate may well change over the course of follow-up, as illustrated in Figure 15-1 for the three treatments continued to the end of the CDP. The rate declined with time, but only slightly. The niacin treatment group had the highest 5-year rate. It was also the group that had the largest number of patients with treatment-related complaints (Coronary Drug Project Research Group, 1975).

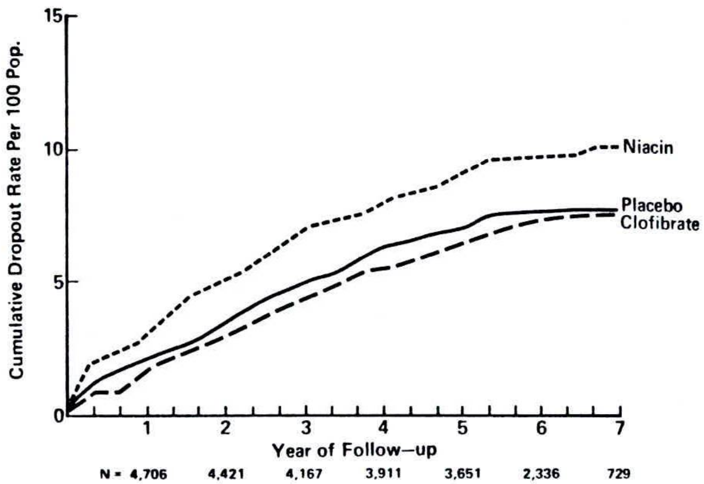
The procedures carried out in conjunction with follow-up examinations may influence dropout patterns. For example, a spurt in dropouts may occur just before an examination involving a noxious procedure. Similarly, there may be a peak after patients pass a specified time point in the trial, especially if they perceive that their time commitments to the trial are satisfied.

A certain number of dropouts in long-term trials will occur simply because of patient relocations. Such losses can be reduced in multicenter trials by transfer of follow-up responsibilities to sister clinics. The CDP was able to maintain the clinic visit schedule for several patients in this way (Coronary Drug Project Research Group, 1973a).

Dropouts should be contacted at periodic intervals. The contacts may be made via home visits, telephone, or mail and should be made even if they cannot be used to collect outcome data since they may be useful in persuading patients to return to active follow-up.

Patients who cannot be contacted should be traced so that contact may be re-established. The tracing process should be initiated as soon as possible. Table 15-3 provides a list of some of the methods that can be used for tracing (see Section 12.5.12 for a discussion of the types of identifying and locator information that should

Figure 15-1 Lifetable cumulative dropout rates for the clofibrate, niacin, and placebo treatments in the CDP.



Note: N denotes total number of patients in clofibrate, niacin, and placebo groups combined. Approximate numbers for individual treatment groups are 2/9, 2/9, and 5/9 times N for clofibrate, niacin, and placebo, respectively.

Source: Reference citation 107. Adapted with permission of the American Medical Association, Chicago, Ill (copyright © 1975).

be collected). Simple steps (Part A, Table 15-3), such as those involved in checking phone and address directories, may enable clinic staff to locate most of the "lost" patients, and they should be carried out before any of the approaches listed in Part B of Table 15-3 are considered.

Searches carried out by agencies retained for that purpose should be done discreetly, without patient contact. This proscription should extend to the coordinating center or other resource centers in the trial as well, unless the patient has had prior contact with the center in question or consented to such contact when he was enrolled.

The cost of searches carried out by firms, such as Equifax,[1] will vary from a few dollars to several hundred, depending on the extent of the search. Relatively inexpensive searches may locate the majority of lost patients, whereas a fairly large investment may be needed to locate those that are especially hard to find. Some help in the location process may be provided by governmental agencies. As a rule, they will not release address information but they may reveal whether their record indicates that the patient has died or may agree to send letters to study patients that are alive, suggesting that they recontact the study clinic.

1. Equifax is an Atlanta-based firm that was established to provide credit and related information for clients of the banking and insurance industry. A branch of the firm was established in the 1970s for marketing a locator service for follow-up studies.

Table 15-3 Methods for relocating dropouts

A. Ordinary
- Via check for address change through the post office, city directories, telephone books, etc.
- Via contact with known friends or relatives of the patient
- Via other sources, such as the patient's most recent employer, church group, etc.

B. Special
- Via a private agency specializing in locating people
- Via firms maintaining large address files and that market a tracing or follow-up service
- Via departments of motor vehicles*
- Via a government agency, such as the Social Security or Veterans Administration*
- Via a private or public institution, such as a hospital*
- Via the patient's private doctor*

*May not yield direct contact with patient if the agency or individual is unwilling to supply the desired address information or is legally constrained from doing so.

Contact with the patient or his family is essential for most forms of follow-up. One notable exception is for mortality follow-up using the National Death Index—NDI (National Center for Health Statistics, 1981). Table 15-4 lists the items of information needed for such searches. The Index contains deaths recorded in the U.S. since 1979. It contains basic identifying information for each deceased person, including the death certificate number and state in which the certificate is located.

It should be possible, with the search methods described above, to provide mortality data on virtually every patient enrolled in a trial. Both the CDP and UGDP were able to achieve this goal (without the NDI since it was not operational when these studies were done). The CDP had vital status on all but a few of the 5,011 patients covered in the final report on clofibrate and niacin (Coronary Drug Project Research Group, 1975). The 1970 publication from the UGDP on tolbutamide provided mortality data on all but 5 of the 823 patients included in that report (University Group Diabetes Program Research Group, 1970e).

15.4 CLOSE-OUT OF PATIENT FOLLOW-UP

The process of disengaging patients from a trial may require as much skill and care as the enrollment process. Recent papers have addressed aspects of the close-out process (e.g., Hawkins and Canner, 1978; Klimt and Canner, 1979; Klimt, 1981). Table 15-5 provides a summary list of considerations that should be addressed in planning for close-out. (See Chapter 3 and Appendix D for additional information.)

Table 15-4 Data items that may be used in searches of the National Death Index

- Last name*
- First name*
- Middle initial

- Social security number*
- Month,* day, and year* of birth
- Father's surname* (for females)

- Age at death (actual or estimated)
- Sex
- Race

- Marital status
- State of residence
- State of birth

*Considered to be key in checking for a possible record match

Table 15-5 Study close-out considerations

- Time schedule (i.e., whether to close-out follow-up for all patients at the same calendar time or after a fixed period of follow-up, see Section 11.7)
- Information to be collected (see Section 15.4)
- Phased treatment disengagement (usually applicable only to drug trials, see Section 15.4)
- Nature of recommendations given to patients regarding subsequent treatment
- Method for ensuring proper transfer of patient care responsibilities to alternate clinic or physician when appropriate
- Ensuring patients have ample opportunity to make alternative arrangements for care and to have any questions answered regarding the trial and its outcome before separation
- Method of summarizing baseline and follow-up data for subsequent use by patient's private physician
- Nature of patient contact required to document separation from trial
- Update of patient locator information and consent (applicable only if there is any possibility of having to contact patients later on to check their status or to recall them for examination)
- Masked trials: Time at which treatment is to be unmasked for study staff; for patients
- Masked trials: Amount of information to be collected on the efficacy of the mask (see Section 15.4 and Section 8.5)

The separation can be an emotional experience for both patients and clinic staff. It should be based on a detailed plan that has been constructed and reviewed before the start of close-out. The details of the separation should be discussed with patients well before separation occurs. Clinic staff must spend whatever time is necessary to answer questions and to help find suitable alternative sources of care. The latter step is imperative in any trial that has been providing patients with routine medical care, as in the UGDP. Investigators in that study discussed care requirements with each patient before departure and made certain of continued care after the close of the trial. The study clinic provided the new clinic or physician with a summary (prepared by the coordinating center) of key baseline and follow-up data assembled on the patient when transfers of care were involved.

The record generated in conjunction with separation should contain:

- The name of the treatment the patient was on
- The date the patient was informed of the treatment assignment (for masked trials)

- The date treatment was discontinued (when appropriate)
- The date of the final close-out visit
- The name of the clinic or physician responsible for future care of the patient
- The treatment recommendation and prescription (when appropriate)
- A list of materials and information given to the patient on departure

Close-out provides an opportunity to assess the adequacy of the mask in masked trials. Theoretically, such checks could be made at various time points throughout the trial. However, usually they are not carried out because of a desire to discourage speculation concerning the treatment assignments since the assessment involves asking the masked individual(s) to state a guess regarding treatment assignment (see also Section 8.5 and Krol, 1983).

A key consideration at close-out has to do with whether to carry out added data collection on patients as they are separated from the trial (the same consideration may arise in conjunction with protocol changes involving termination of a particular treatment during the trial). The wisdom of making such provisions depends on the importance of the data generated in relation to the aims of the trial. Results obtained for tests or procedures for which there are no corresponding baseline values will be of limited use in making treatment comparisons if the treatment groups differ because of losses due to dropouts or deaths. Investigators in the CDP opted against introduction of special data collection schemes during close-out, except for the addition of a few items to facilitate relocation of patients (Krol, 1983).

The method of terminating therapy in masked drug trials must be given special consideration. A dosage step-down scheme may be necessary if an abrupt cessation of one or more of the drugs is considered unsafe. In addition, a patient will want to know the treatment he was on. Hence, study physicians must be supplied with treatment codes well in advance of close-out visits, especially in trials where time is needed to consider alternative courses of therapy before making a treatment recommendation.

Ideally, any treatment recommendation given to patients at close-out should be based on findings from the trial. However, often this is not possible, since the final analysis of the results may not be completed by the time of close-out. Recommendations may have to be qualified or simply withheld, especially in designs involving close-out after a fixed period of follow-up (see Section 11.7). In such cases, the close-out process will extend over a period of time as long as that required for patient enrollment. It may not be advisable to unmask treatment assignments in such designs until all patients have been separated from the trial, unless it is possible to lift the mask on a per-patient basis (see Section 10.5).

Patients should be told at close-out if the clinic plans to keep in touch with them and, if so, the reason for doing so and the way in which contact will be maintained (e.g., via mail, telephone, or home visits). They should be asked to sign a consent authorizing the contacts and to provide updated locator information if contacts are planned. In fact, it is a good idea to alert patients to the possibility of future contacts and to obtain consents for them even if subsequent contacts are not planned, if there is any chance they will be needed later on.

15.5 TERMINATION STAGE

Close-out of patient follow-up is only the first stage in shutting down the trial. It is normally followed by a series of activities (see Table 15-6 and Section VI of Appendix D) beginning with completion of the close-out visits and ending with termination of all funding for the trial. The time needed for termination is variable and depends on the trial. A period of a year or longer is common for trials of the type sketched in Appendix B.

As a rule, clinics will require financial support for a period of time beyond the patient close-out stage to complete data transmissions to the data center and to respond to edit queries from that center. Support for the data center will have to extend beyond that for clinics to allow adequate time for the center to complete analyses of the results and to prepare them for publication. The UGDP Coordinating Center continued to receive funding through April 1982, nearly 7 years after completion of the last close-out examinations. The coordinating center in the CDP continued to operate through 1983, over 9 years after termination of the closeout stage of that trial.

One of the last steps in the termination stage has to do with record storage and disposition. All study forms and related documents to be retained (especially those with personal identifiers on them) should be stored in a secure location. Forms and related documents should be

Table 15-6 Activities in the termination stage

A. General

- Revise organizational structure (at the start of the termination stage) to meet special needs of the termination stage. Discharge committees no longer needed
- Update mortality follow-up for all patients, including dropouts
- Carry out final data edit checks
- Establish cutoff date beyond which changes to the data system are no longer allowed (needed so data files can be "frozen" for final analysis)
- Develop and implement plan for the final disposition of the study data forms and related documents, such as x-rays, fundus photographs, ECGs, etc.
- Develop plan for dealing with requests for special analyses or for access to the study data after termination of study funding (see Chapter 24)
- Disseminate study findings and conclusions to study investigators and to referral physicians (may be done by distributing preprint or reprint of main study manuscript)
- Discharge all remaining committees at the end of the termination stage

B. Additional activities in drug trials

- Collect sample of study drugs for future laboratory analysis in case of questions regarding drug purity
- Dispose of remaining unused study drugs
- Submit final report to the FDA if trial involved an INDA or IDEA; cancel INDA or IDEA after acceptance of the report by the FDA

destroyed (in compliance with local statutes for disposition of medical records) if secure storage cannot be assured and the required period of storage has passed (see Section 17.6). General factors to consider in arranging for record storage and policy questions concerning access to study data are discussed in Chapter 24.

15.6 POST-TRIAL PATIENT FOLLOW-UP

Post-trial follow-up, by definition, takes place after the termination stage of the trial (see Chapter 3, Appendix D, and Glossary for further details). Ideally, the patient personal identifiers needed for the follow-up should be deposited at a central location before the trial terminates, especially in multicenter trials. If this is not done, the task of assemblying the information after the trial has terminated may make subsequent follow-up difficult, if not impossible. The repository should be established at a center that can assure secure storage, and that is likely to remain functional into the foreseeable future. Federal agencies, such as the National Institutes of Health (NIH), generally are not suitable as a repository because of their susceptibility to requests under the Freedom of Information Act (see Chapter 24).

There should be a sound rationale for any post-trial follow-up involving direct patient contact. The prime motivation for most post-trial follow-ups stems from a desire to extend the period of observation for death or some other serious but nonfatal event. Another reason may be to observe patients for a disease or condition that may be caused or aggravated by treatments administered during the trial. The usefulness of the information obtained will depend on the completeness of the follow-up and the nature of intervening treatments administered after close-out. Interpretation of the results will be easiest if patients have not been exposed to any additional treatment after separation from the trial. It will be problematic if they have been.

The CDP provides an example of post-trial mortality follow-up. The follow-up was performed by the coordinating center, with help from clinics still in operation when the follow-up started in 1981. Addresses and other identifying information on patients were used for tracing them and for accessing the National Death Index and other files.

Some trials have provided a form of post-trial follow-up during the trial. For example, this was the case for the two discontinued treatments in the UGDP. Patients assigned to both the tolbutamide and phenformin treatments were followed for mortality (as well as for other nonfatal events) until separation of all patients in August of 1975. There has been no further post-trial follow-up of any of the UGDP treatment groups since then (University Group Diabetes Program Research Group, 1982).

16. Quality assurance

If it ain't broke, don't fix it.

Old American Adage

16.1 INTRODUCTION

Quality assurance, as applied to clinical trials, is any method or procedure for collecting, processing, or analyzing study data that is aimed at maintaining or enhancing their reliability or validity. Examples include (see Table 16-1):

- Edit procedures to check on the accuracy of items on completed data forms
- Repeat of a laboratory determination to check on reproducibility
- Rekeying data as a check for errors in the entry process
- Carrying out analyses by clinic in a multicenter trial to detect performance variations
- Reprogramming an analysis procedure as a means of checking on its accuracy

Deficiencies anywhere in the chain of events from data generation to publication of the results can reduce the quality of the finished product and the conclusions reached from the trial. Everyone involved in data collection, analysis, and manuscript writing must perform effectively to produce a quality end result.

This chapter deals with the mechanics of quality assurance. Other chapters of this book touch upon issues related to quality assurance. They include:

- Treatment masking (Chapter 8)
- Randomization (Chapters 8 and 10)
- Data form construction (Chapter 12 and Appendix F)
- Production and maintenance of study handbooks and manuals (Chapter 13)
- Testing the data intake and processing system (Chapter 13)
- Database maintenance (Chapter 17)
- Review procedures for study publications (Chapter 24)
- Activities staging (Appendix D)

16.2 ONGOING DATA INTAKE: AN ESSENTIAL PREREQUISITE FOR QUALITY ASSURANCE

Most of the quality assurance procedures outlined in Table 16-1 require a continuous and timely flow of data from the clinic to the data center to be useful. The data edits and analyses carried out during the trial to assess data quality and clinic performance will lose much of their value if there is a large time gap between data generation and conversion into computer-readable formats.

The ideal data intake system is one in which data are edited and entered on the day of generation, or very shortly thereafter. Theoretically, the entry process could take place as patients are examined using video displays to remind physicians and technicians of items to be entered. However, on-line data entry of this sort is usually not practical. The need to do so during an examination may distract both the patient and physician and may complicate the examination. Further, it is unlikely that all data could be entered on the spot since much of it may not be available until some time after the examination is completed (e.g., as with results of certain laboratory tests or readings from biopsy material, ECGs, X rays, etc.). However, even if these problems could be overcome, documentation of the data collection process argues against on-line entry. The data forms and related paper records are needed to document the data collection and entry processes, to say nothing of their use in patient care. Hence, discussion throughout this book is predicated on the assumption that data collection always involves completion of paper forms and records, regardless of where and how data entry is done.

Table 16-1 Quality assurance procedures

- Visual check by a member of the clinic staff after a data form is completed for illegible responses and for unanswered or incorrectly answered items
- Ongoing data processing
- Replication of the coding and data entry process as a means of error detection
- Computer edit of keyed data for inadmissible codes or missing values
- Data edit queries (directed from the data center to the clinic) concerning completed data forms
- Generation of periodic status reports concerning the data collection process
- Repeat laboratory determinations
- Multiple independent readings of ECGs, fundus photographs, X rays, tissue slides, etc.
- Independent review of patient death records for classifying cause of death
- Submission of masked duplicate specimens or records to check on the reproducibility of a measurement or reading procedure
- Generation of periodic reports assessing the compliance of clinics to the treatment protocol
- Comparison of the performance of clinics in a multicenter trial to detect differences in the quality or completeness of the data generated, as reflected by such characteristics as number of missed follow-up examinations, number of dropouts, number of deficient data forms, etc.
- Reprogramming of a data editing or analysis procedure as a check on program accuracy or on the quality of program documentation
- Interim analyses of study data for treatment effects that can be used to reveal inadequacies or inconsistencies in the data collected

One viable approach to on-site data entry involves completion of a paper form during the patient examination and then entry of the information contained on that form as soon after the examination as possible—ideally, on the same day or within a few days after the examination. The entry should be done by clinic personnel who are familiar with the data collection requirements of the trial, and should be subjected to edits during the entry process. The keyed data may remain at the clinic for subsequent analyses or may be transferred to a central data facility for additional edits, analysis, and storage. The transfer may take place on-line as the data are keyed, or may be done off-line either on a fixed schedule or on demand, as dictated by the data center. On-line transfer may be via hard-wired or telephone connections to the central facility. Off-line transfers may be done by telephone or by mailing the magnetic records to the central storage facility.

Systems involving on-site data entry and multiple data generation sites, as in multicenter trials, are herein referred to as distributed data entry systems. Those in which data forms are sent to a data center for entry are referred to as centralized data entry systems. All but two of the 14 trials sketched in Appendix B had systems of the latter type. Only the Coronary Artery Surgery Study (CASS) and Hypertension Prevention Trial (HPT) had distributed data entry systems.

A trial in which each data generation site is responsible for maintaining its own database with programs provided from the data center is herein referred to as having a distributed database (e.g., the HPT). A trial in which the only electronic database that exists is the one maintained at the central data facility is herein referred to as having a centralized database.

The main advantages of distributed data entry have to do with the potential for eliminating the time lag between the data generation and data entry processes, and with the ability to involve data collection personnel in the data entry process. However, in order to work well, the approach requires skilled personnel at the data center who have the patience and know-how to select the equipment needed for the system, to supervise acquisition and installation of it at the clinic, to train clinic personnel in its operation, and to develop and maintain the software packages needed for on-site data entry and editing.

The lag time between the generation of a form and data entry should never be more than a week or two, regardless of the type of entry system. The goal should be to establish and maintain the discipline needed to ensure a timely flow of forms from the point of origin through data entry. Designs that allow forms to accumulate over a specified time interval or in batches of a certain size before they are forwarded for data entry should be avoided. The best design is one in which individual forms proceed to data entry on a per-form basis without regard to other forms or conditions. Batching increases the time from completion of a form to data entry. If some batching is required for reasons of efficiency, it should be minimal and should never allow forms to accumulate for more than a week or two. The same is true for accumulation of forms at the data entry site.

16.3 DATA EDITING

The term *data editing* refers to the process of detecting, querying, and, when appropriate, correcting values in a data set that are invalid. The normal editing process involves a series of edit checks and edit queries. An edit check is an operation carried out on an item or series of items on a completed data form for the purpose of identifying possible errors (see Table 16-2). An edit query is a question generated from review of a completed data form that concerns the accuracy or adequacy of some item of information on the form and that requires someone at the generation site to review the information in order to respond to the query. The query may be generated by a clerk checking a completed form for deficiencies, or by a CRT or printer driven by edit programs.

Edit queries that are written will be referred to as edit messages. Any edit message that requires review and possible corrective action should be printed on hard copy. This does not preclude use of a CRT for a preliminary display of messages, but this procedure is not adequate if messages must be sent to various places in the clinic for review and action. Special care must be taken to make certain that the messages are intelligible. Table 16-3 gives suggested edit message rules.

A sample of such messages, as taken from the Macular Photocoagulation Study (MPS), is reproduced in Figures 16-1 and 16-2 for a fictitious clinic and patient. The two pages relate to

Table 16-2 Types of edit checks

Type	*Edit check*
• Patient identification and record linkage	• Check of ID number and name code for transposition errors
	• Check of name for spelling errors
	• Check to make certain all pages of a given form carry the same ID number
• Legibility	• Check for illegible handwritten replies, spelling errors, etc.
	• Check for response checkmarks placed outside designated spaces
• Form admissibility	• Check to determine if the form was completed within the specified time window
	• Check to make certain the form completed is the correct one for the indicated examination
• Missing information	• Check for unanswered items or sections of an otherwise completed form
	• Check to make certain all required forms have been completed
• Consistency	• Check of information supplied in one section against another section on the same form for inconsistencies
	• Check of information supplied on the same patient on one data form with that from another form completed at the same or at a different examination as a check for possible data inconsistencies
• Range and inadmissible codes	• Check to identify items with values that exceed specified ranges
	• Check for undefined alphabetic or numerical codes

Table 16-3 Edit message rules

- Use a format that facilitates use by clinic personnel, even if the format is not ideal for data entry
- Test the intelligibility of the messages on personnel who must deal with the messages
- Avoid the use of esoteric codes, abbreviations, and other symbols that are not readily understood by personnel who must respond to the statements
- Identify the patient, examination, form, and item number on the edit statement
- Allow space on the statement for the respondent to indicate the action taken
- Group messages for a given patient examination in such a way so as to simplify the task of dealing with them (e.g., list all laboratory-related edit messages for a given examination on one page and all messages concerning clinical evaluation of the patient on another page, if different personnel are required to deal with the two types of edit messages)
- Generate duplicate copies of the edit messages to allow clinics to retain a copy of answered queries

patient 03-072-S, with name code MARV, seen on July 6, 1983, in connection with his fifth follow-up clinic visit (second annual examination). The message dated August 3, 1983, relates to inconsistencies noted in visual acuity measurements done on the patient. The message dated October 4, 1983, relates to discrepancies in readings of fundus photographs done at the clinic with those done at the MPS Fundus Photography Reading Center. Clinic personnel are required to indicate corrected values on the edit message sheets and to return them to the MPS Coordinating Center for processing.

The first set of edit checks should be done by hand at the clinic shortly after a form is completed. A second set of checks, involving a combination of hand and computer checks, may be performed when the data are keyed. The main advantages of computer checks lie in the ease and accuracy with which they can be made and in the ability to use the computer to write and

Figure 16-1 MPS Coordinating Center edit message of August 3, 1983.

```
Clinic : 03   Eye Research Clinic                              Study:  SMD
Patient :  03-072-S      Code : MARV                           Visit:  fv05    07/06/83
                                                                  (Follow-up Visit 05)

    ** COMPONENT   0702   Visual Acuity Measures (Follow-up)

          ITEM                    OLD VALUE                 CORRECTED VALUE
          ====                    =========                 ===============

          4AR                     10                        --
          4BR                     99                        --
          4CR                     00                        --

    There is a problem with one or more of the above answers.  Question 4AR
    must be answered with either a '10' or a '05', and the answers to Questions 4BR
    and 4CR must indicate the smallest line read at THAT distance and the number
    of additional letters read at THAT distance.
    Please supply the correct answers for all three questions.



                              PERSON COMPLETING THIS FORM: ____________________ DATE: ________
```

Figure 16-2 MPS Coordinating Center edit message of October 4, 1983.

```
*** MPS READING CENTER ***

Clinic : 03   Eye Research Clinic                    Study:  SMD

Patient :  03-072-S    Code : MARV                   Visit:  fv05   07/06/83
                                                        (Follow-up Visit 05)

   ** COMPONENT   5511   Annual Follow-up Grading Form (PT, FV01, FV03 ...)

          ITEM              OLD VALUE             CORRECTED VALUE
          ====              =========             ===============


          1a                n                     _

          5b                5                     _

          6b                n                     _

     If there is no blood, there is no blood !
     In other words, if question 1a is 'n' then questions 5b and 6b must also be 'n'.



          10a               9                     _

          10b               n                     _

     If photos or FA are present then questions 10a and 10b  MUST  be answered.
     If there is no RPE atrophy (Question 10a = 'n') then question 10b must also be 'n'.



                PERSON COMPLETING THIS FORM: ______________________ DATE: ________
```

keep track of the queries. Clinics need periodic reminders of outstanding queries to ensure they are addressed (see Chapter 17 for a discussion of file updates based on edit changes). The computer, however, should never be a substitute for the checks performed by staff at the clinic before forms are forwarded for data entry. An experienced clinic coordinator, with an eye for errors and an encyclopedic knowledge of the study protocol, can do more to enhance the quality of the data generated than any set of computer checks.

There should be an audit trail for any change made to a completed data form, regardless of when and how the change was initiated. The nature of the deficiency, when it was detected, the change, and when the change was made should be noted. Once recorded on a form, data should not be erased or obliterated. Entries that are incorrect should be lined out and the new entries added to the form. Any change, regardless of when it was made, should be dated and should carry the initials of the person making the change.

Data entry personnel should be given explicit instructions regarding the types of data changes they may make. Sound practice dictates that data should be entered as recorded, even if an item is "clearly" in error and the change required seems obvious. The temptation is to make an "obvious" change on the spot, without any checking. However, there are at least two reasons to resist the temptation. First, there is always the chance that the item has been correctly recorded even though it appears to be in error. Second, on-the-spot changes will lead to discrepancies between the computer data file and the original study records. Such discrepancies, if sizable, may lead to serious questions concerning the integrity of the data collection and processing activity. Both audits of the University Group Diabetes Program (UGDP) focused on the accuracy of the data collection and entry processes, as evidenced by comparisons of values recorded on the original study forms with entries appearing in the computer data file of the study (Committee for the Assessment of Biometric Aspects of Controlled Trials of Hypoglycemic Agents, 1975; Food and Drug Administration, 1978). Fortunately, procedures in the UGDP Coordinating Center required all changes to the computer file to originate with the original data forms. It would not have been possible to maintain a one-to-one correspondence between the original records and computer file without such a rule.

A series of identification and linkage checks should be performed before any form is added to the computer file. The ID number recorded should be checked for transposition errors (e.g., via a check digit; see Glossary). No form should be added to the file unless the ID number and other identifiers agree (e.g., such as name or name code).

Admission of a record to the data file may also depend on time window (see Glossary) checks needed to ensure that the information in question was obtained within a specified time interval. Examinations performed outside the specified window may either be rejected or assigned to the appropriate time slot, depending on the philosophy of the study.

Computer checks made during data entry should be designed to detect use of inadmissible codes (e.g., entry of an alphabetic character when only numeric codes are permissible or use of an undefined or inadmissible numeric code). These errors should be corrected before the generation of edit messages.

Most editing systems are designed to deal with one item at a time. There may be some cross checking of items, but it is usually limited to items on the same form. Cross checking of items across forms is generally not done because of the logistical difficulties involved in making such checks and because of the limited return in added undetected errors and deficiencies.

The foundations for data editing should be laid when the study is designed. The edit requirements should be specified in the handbooks and manuals needed for operation of the trial. The data forms used in the trial, as suggested in Chapter 12, should include reminder and documentation items (see Section 12.5.13) that require clinic personnel to carry out essential checks while the forms are being completed and that remind them of the steps that must be performed in conjunction with specified data collection procedures.

16.4 REPLICATION AS A QUALITY CONTROL MEASURE

Replication of an observation or reading is frequently used as a check on the quality of the data obtained. Examples of replication used in this way are:

- Comparison of two independent measurements, such as a laboratory test, to determine if the difference observed is outside a specified range
- Use of two independent readings of an ECG to identify items of disagreement for adjudication by a third reader
- Comparison of cause of death codes assigned by two different individuals to identify areas of disagreement for adjudication by a third reader
- Averaging two or more consecutive blood pressure readings made on a patient during a given clinic visit in order to have a more reliable estimate of the patient's "true" blood pressure
- Rekeying data as a check for errors in the entry process
- Use of a computer program, written specifically to duplicate the tasks performed by another program, to check the accuracy of results provided with the original program

Replicate values obtained from repeat readings or from aliquots of the same laboratory specimen are usually combined by averaging to yield a single composite value. However, this approach is not suitable for combining independent readings made from the same record involving binary measures (e.g., presence or absence of S-T depression on ECGs). Some form of adjudication is necessary when the readings disagree. It may be done by having an "expert" make the judgment or by asking the individual readers to reach agreement. It is important to select readers who work well together and who interact on a peer basis if the latter approach is used.

A common problem in trials involving laboratory determinations has to do with the detection and disposition of outlier values (see Glossary). Explicit rules are required to indicate the conditions under which a determination is to be repeated and the value or values to be reported in such cases. The procedures of the laboratory performing the determinations should be reviewed when the rules are constructed. Laboratories differ with regard to the practices they follow in making repeat determinations because of suspected errors. Some of those practices can bias the results reported, for example, as is the case with a laboratory that does three determinations per sample, but reports only the two most concordant values. The same is true for a laboratory that opts to make repeat determinations when the observed inter-aliquot difference for the first set of determinations exceeds a prespecified limit and then reports only the results of the second set.

The easiest, and often the best, rule to follow is one that requires the laboratory to report all determinations made, without any censoring. Outlier values which, if retained in the data file, would have undue influence on means and variances may be eliminated or trimmed when the analysis tape is written (see Section 17.7).

16.5 MONITORING FOR SECULAR TRENDS

A secular trend in the readings made from records, such as ECGs, X rays, fundus photographs, and the like, or from laboratory determinations, can be troublesome, especially if differential by treatment. The possibility of this happening is minimized when the ordering of the readings or determinations is independent of treatment assignment (e.g., in schemes in which readings or determinations are done on an ongoing basis and in the order of generation). However, even so, it is wise to monitor for trends. The information is useful in characterizing the magnitude of the trend and in indicating whether it is differential by treatment. Assurance in the latter regard is especially important for readings or determinations that are not masked with regard to treatment assignment or that are ordered by treatment assignment. In addition, characterization of the trend, even if not needed for making treatment comparisons, is useful when evaluating follow-up results for a particular treatment group in natural history studies.

The number of repeat determinations or readings that are made should be dictated by the importance attached to detecting time trends and the total resources available for quality control. The cost of maintaining systems designed to detect secular trends can be sizable. Only a small part of the cost may be associated with making the actual readings or determinations. The larger costs will be associated with managing the monitoring system.

Monitoring a laboratory or reading center for secular trends requires use of known standards that are subjected to repeat analyses or readings over the course of the trial. To be useful, the repeat specimens should be indistinguishable from other specimens received at the laboratory or reading center.

Developing a reliable set of standards, at least for laboratory determinations, is not a trivial task. The problem would be easily solved if a single set of standards could be used throughout the trial. However, most biological substances degrade with time and, hence, more dynamic approaches are needed. The Coronary Drug Project (CDP) created a pool of donor serum. The pool was aliquoted and then frozen (Canner et al., 1983c; Hainline et al., 1983). Specimens from the pool were submitted to the central laboratory on a time schedule designed to coincide with actual patients in the trial, using ID numbers of deceased patients.[1] When a given pool was near depletion, or the time limit set for its use was about to expire, a new one was created. Use of specimens from the new pool overlapped use of specimens from the old pool, so as to provide a basis for estimating concentration differences between the two pools.

Similar monitoring is needed for readings of ECGs, fundus photographs, X rays, biopsy materials, etc. However, the mechanics of setting up and maintaining systems for this purpose are even more complicated than those required for laboratory determinations. The CDP used a system for making repeat ECG readings to monitor for time-related shifts in reading standards (Coronary Drug Project Research Group, 1973a). However, the system was difficult to manage, and it was not easy to keep readers from identifying repeat tracings, especially those with distinctive patterns. In any case, the system was only effective in detecting short-term trends since the tracings chosen for rereading were selected from batches of tracings that had been read in the recent past. Inclusion of records read in the distant past was not practical because of date information contained on the tracings. There was concern that lack of homogeneity of

1. Use of fictitious ID numbers would have caused the central laboratory to reject the specimens because of edit checks performed by it prior to admitting specimens for analysis.

dates within a reading batch would enable readers to identify repeat tracings.

Another method sometimes used to control a reading process involves use of reference measurements or records to help readers gauge the degree of abnormality seen in actual records. This approach was used in the MPS for grading the severity of certain kinds of eye abnormalities, as seen in fundus photographs. The severity of an observed abnormality was graded by selecting the photograph from an ordered set of standard photographs that was most similar to the one in question.

Concerns regarding secular trends are obviated if records are read over a short period of time at the end of the study and in a random order with regard to the time of generation and treatment assignment. However, this approach suffers from two major disadvantages. First, postponing readings until the end of data collection means that results from the records in question will not be available for interim analyses during the trial (see Chapter 20). Second, waiting for the readings may delay preparation of the final report. Both disadvantages are avoided with an ongoing reading program that runs over the course of the trial.

16.6 DATA INTEGRITY AND ASSURANCE PROCEDURES

An editorial by Meinert (1980b) discusses factors that may contribute to dishonest practices in the field of clinical trials. They do occur, but there is no reason to believe their incidence is higher in this field than in other areas of research. In fact, it may be lower because of the general emphasis on error detection and quality control. However, even so, there are good reasons for constant vigilance against shady practices. The luxury of replication, used so effectively in the laboratory sciences to confirm or refute findings, is not always feasible in clinical trials for practical as well as ethical reasons. For example, it would be difficult to justify additional placebo-controlled trials of hypertensives in the light of the conclusions from those done by the VA (Veterans Administration Cooperative Study Group on Antihypertensive Agents, 1967, 1970) or to replicate the Multiple Risk Factor Intervention Trial in view of its cost and the time required to complete it (Multiple Risk Factor Intervention Trial Research Group, 1982).

Table 16–4 Data integrity checks

- Comparison of information on a patient's medical chart with that recorded on a study data form
- Comparison of information on data forms with that in the computer
- Interviews with support personnel for identification of questionable or undesirable data practices
- Review of methods for issuing treatment allocations to check for discrepancies in the administration of the allocation schedule
- Review of analysis procedures used by the data center for evidence of a bias for or against a particular treatment
- Comparison of the distribution of inter-aliquot differences to detect clinic differences in reading or reporting procedures
- Independent audit of published reports to determine if the conclusions are supported by the raw data.

Table 16–4 provides a list of checks that can be performed to help identify questionable data practices, whether due to honest errors, careless oversights, or purposeful acts. The checks, like others in the trial, should be ongoing since the problems they are aimed at detecting can occur at any time over its course.

The best preventive measure is a staff that appreciates the importance of honesty and integrity in all aspects of the trial. The responsibility for instilling the proper philosophy rests with the leaders of the trial. They must, by the statements they make and the actions they take, set a tone and standard that permeates the entire investigative group.

16.7 PERFORMANCE MONITORING REPORTS

It is good practice to prepare reports summarizing performance characteristics of the trial as it proceeds. The reports should be prepared by the data center and should be designed to provide up-to-date information on all relevant activities of the trial. Some of the performance characteristics that should be monitored are listed in Table 16-5. See also Appendix G for sample reports.

The information in the report should be reviewed by the leadership of the trial (e.g., the steering committee) and should be used as a basis for initiating corrective action, where appropriate. To be useful as a monitoring tool, reports should indicate the relative standings of

Table 16-5 Performance characteristics subject to ongoing monitoring

A. Clinic characteristics

1. Patient recruitment

- Number of patients screened for enrollment; proportion rejected and tabulation of reasons for rejection*
- Current rate of recruitment compared with that required to achieve a prestated recruitment goal

2. Patient follow-up

- Distribution of enrollment times and median length of follow-up
- Number of completed follow-up examinations*
- Number of missed examinations and number past due*
- Rate of missed examinations*
- Number of dropouts*
- Total number of dropouts and estimated dropout rate
- Number of patients who cannot be located for follow-up

3. Data quantity and quality

- Number of forms completed since last report and number that generated edit messages
- Current edit message rate per form contrasted with rates from previous time periods
- Number of forms received with missing parts or missing supporting records
- Number of unanswered edit queries*
- Number of patients enrolled with incomplete baseline information*

4. Protocol adherence

- Number of ineligible patients enrolled*
- Number of patients who did not accept the assigned treatment*
- Number of patients who received a treatment other than the one assigned*
- Summary of data on pill counts and other adherence tests by treatment group*
- Number of departures from the treatment protocol*
- Summary of other treatment or data collection protocol violations

B. Data center characteristics

- Number of allocations issued*
- Number of allocations returned unused
- Number of forms received*
- Total number of forms awaiting data entry*
- List of coding and protocol changes implemented since last report
- List of data processing and programming errors and likely impact on study results
- Summary of major events, such as computing malfunctions, necessitating use of backup tapes to restore the data system
- Timetable for unfinished tasks

C. Central laboratory characteristics

- Number of samples received*
- Number of samples received improperly or inadequately identified*
- Number of samples lost or destroyed*
- Number of samples requiring reanalysis and tabulation of reasons for reanalysis*
- Backlog of samples remaining to be analyzed*
- Summary of major events affecting laboratory operations, such as power outages, particularly those resulting in possible degradation of frozen samples
- Mean and variance of inter-aliquot differences over time for specified tests
- Secular trend analyses based on repeat determinations of known standards

D. Reading center characteristics

- Number of records received and read*
- Number of records received that were improperly labelled or had other deficiencies (tabulate deficiencies)*
- Analyses of repeat readings as a check on reproducibility of readings and as a means of monitoring for time shifts in the reading process

E. Other performance characteristics

- Status of papers being written
- Progress in locating patients lost to follow-up
- Labelling errors made in drugs dispensed from the central pharmacy

*Report should contain results for the entire study period, for the time period covered since production of the last report, and for the last one or two preceding time periods.

clinics in multicenter trials with regard to important functions such as patient recruitment, completeness of follow-up, number of error-free forms, etc. The tabulations may be for the entire study period or for defined time intervals (e.g., the last 3 months, 4 to 6 months ago, etc.). The rankings can be helpful in identifying problem clinics. However, they should be viewed with caution when used as a basis for taking corrective or punitive action involving individual clinics. The range of the difference between the best and worst clinics with regard to a performance

statistic is more important than clinic rankings.

Members of the entire investigative group should have access to the performance monitoring reports to enable them to gauge their standing in the study. Peer pressure, exerted via dissemination of the information, can be helpful in encouraging clinics with poor performance records to improve.

16.8 OTHER QUALITY CONTROL PROCEDURES

16.8.1 Site visits

A site visit, used in this context, is:

A visit to a center in a trial made by personnel from outside that center for the purpose of assessing its performance or potential for performance.

Those making the visit may be from other centers in the trial or from outside the trial. The size of the visiting team will be dictated by the nature of the visit. It may be done by just one person or it may involve a half dozen or more people depending on needs and circumstances. (See Cassel and Ferris, 1984, for details regarding clinic visiting procedures in an ophthalmic study.) The "typical" clinic visit in a multicenter trial may involve the chairman of the study (or his representative), a director of another clinic in the trial, the director of the data coordinating center (or his representative), and the project officer, as well as other selected resource people (e.g., a clinic coordinator if there are problems in the way forms are completed, or a person knowledgeable in laboratory methods if there are problems in this area).

The head of the visiting team should prepare a written report of the visit, based on input from the entire team. It should indicate when the visit took place, who made it, who was seen, the areas of activities reviewed, and the strengths and weaknesses of the center. When appropriate, it should contain a list of specific recommendations. It should be sent to the director of the center visited and to the appropriate leadership body of the study for review (usually the steering committee).

Clinic visits may be made on an as-needed basis or on a set time schedule. The CDP used a combination of the two approaches. The steering committee requested visits of clinics considered to have performance problems. Clinics that were not visited on this basis were visited routinely over the course of the trial.

The visits should include contacts with senior staff as well as essential support staff in the clinic and may involve any or all of the following activities:

- Private meeting of the site visitors with the clinic director
- Meeting of the site visitors with members of the clinic staff
- Inspection of examining and record storage facilities
- Comparison of data contained on selected data forms with those contained in the computer data file
- Review of file of data forms and related records to assess completeness and security against loss or misuse
- Observation of clinic personnel carrying out specified procedures
- Check of handbooks, manuals, forms, and other documents on file at the clinic to assess whether they are up-to-date
- Physical or verbal walk-through of certain procedures (e.g., the series of examinations needed to determine patient eligibility, or the steps followed in the informed consent process
- Conversations with actual study patients during or after enrollment as a check on the informed consent process
- Private conversations with key support personnel to assess their practices and philosophy with regard to data collection
- Private meeting with the clinic director's chief concerning special issues

The visiting process should not be limited to clinics. It should include the data center as well as other key resource centers in a trial. A "typical" data center visit may include many of the activities mentioned above as well as:

- Review of methods for inventorying forms received from clinics
- Review of methods for data entry and verification
- Assessment of the adequacy of methods for filing and storing paper records received from clinics, including the security of the storage area and methods for protecting records against loss or unauthorized use
- Review of available computing resources
- Review of method of randomization and of safeguards to protect against breakdowns

in the randomization process (see Chapter 10)

- Review of data editing procedures
- Review of computer data file structure and methods for maintaining the analysis database
- Review of programming methods both for data management and analysis, including an assessment of program documentation
- Comparison of information contained on original study forms with that in the computer data file
- Review of methods for generating analysis data files and related data reports
- Review of analysis philosophy, especially in relation to the principles discussed in Chapter 18
- Review of methods for backing up the main data file
- Review of methods for restoring the main data file or original study records if lost or destroyed
- Review of master file of key study documents, such as handbooks, manuals, data forms, minutes of study committees, etc., for completeness

Some studies, such as the National Cooperative Gallstone Study (NCGS), have gone a step beyond the process outlined above in monitoring data center operations. It established a special monitoring committee, made up of people from outside the study, with first-hand experience in data coordinating center operations to review operations in the center (National Cooperative Gallstone Study Group, 1981a). The committee was responsible for carrying out periodic reviews of the center and for reporting results of those visits to the NCGS Steering Committee and Advisory-Review Committee.

16.8.2 Quality control committees and centers

Certain of the quality control functions in some of the larger-scale multicenter trials may be performed by specifically constituted committees, as already stated above for the NCGS. For example, the CDP had a laboratory committee to review laboratory standards and methods (Coronary Drug Project Research Group, 1973a). The Aspirin Myocardial Infarction Study (AMIS) created a committee that was responsible for monitoring the performance of all centers in the trial, primarily via performance monitoring reports prepared by the AMIS Coordinating Center (Aspirin Myocardial Infarction Study Research Group, 1980b). Various other committees in the structure of a trial will have quality control functions.

A few studies, such as the Persantine Aspirin Reinfarction Study (PARIS), have funded a quality control center (Persantine Aspirin Reinfarction Trial Research Group, 1980a). The function of the PARIS center was to carry out data audits by comparing data from original study forms with those in computer files at the PARIS Coordinating Center. A second function was to check on the accuracy of analyses performed by the Coordinating Center. A third was to serve as a second analysis center for the study, using tapes provided by the PARIS Coordinating Center.

16.8.3 Data audits

A data audit, as used herein, involves an item-by-item comparison of information recorded on an original study form with that contained in the computer file for that form. Such audits, as mentioned in Section 16.3, were carried out by groups reviewing the UGDP, after the study was finished. To be useful as a quality control measure they must be carried out during the trial. Ongoing audits of this sort are especially important in studies with distributed data systems where forms are keyed at the clinic and, hence, may never be sent to the data center, as in the HPT. Clinics in that study are required to forward a random sample of completed data forms to the data coordinating center. Staff at the center compare entries on the forms with those in the data file. Discrepancies are noted for review. A less systematic approach might involve on-the-spot audits carried out during clinic site visits and done by arbitrarily selecting a few forms for comparison with data listings prepared by the data center in conjunction with the visit.

Part IV. Data analysis and interpretation

Chapters in This Part

The four chapters in this Part deal with issues involved in the analysis and interpretation of results from trials. The first chapter details issues concerned with database management. Chapter 18 details general principles to be followed when results are analyzed. It also contains brief descriptions of commonly used methods of analysis for trials involving a binary event as the outcome measure. Chapter 19 contains a list of questions and short answers concerning the design, analysis, and interpretation of clinical trials. Chapter 20 addresses issues involved in treatment monitoring and provides a brief description of some of the analysis approaches used for that purpose.

17. The analysis database

Round numbers are always false.

Samuel Johnson

17.1 INTRODUCTION

This chapter contains a discussion of issues involved in the development and maintenance of the analysis database. The analyses may be for the purposes of quality control (Chapter 16), safety monitoring, (Chapter 20), or for preparation of publications at the end of the trial (Chapters 18 and 25).

The study database, as defined herein, consists of all data contained on official data forms of the study. It includes data from all baseline and follow-up forms, as well as data from laboratory tests and other procedures (e.g., ECGs, fundus photographs, liver biopsies, etc.) that are a required part of the study protocol. It does not include data that are part of a patient's general medical record, except to the extent that such information overlaps that which is needed for the study.

The analysis database is constructed from the study database, and consists of all codified information contained in the latter database. Ideally, there should be a one-to-one correspondence between the paper forms generated from a study and the analysis database. There will be when all entries on study forms are made in codified form. However, this is not always practical, especially if some of the information collected is recorded in narrative form and is not coded.

17.2 CHOICE OF COMPUTING FACILITY

Most trials will require use of electronic files to facilitate analysis of the study results. The choice of the electronic medium (e.g., tape or disk) and facility is not as crucial in a short-term trial as in a long-term one. The choice of the facility will be between a dedicated one, operated by study personnel for the exclusive use of the study, or a general-use facility, operated by someone else and shared with other users, or a combination of the two kinds of facilities. Table 17-1 outlines the pros and cons of the two classes of facilities.

Once the type of facility has been chosen, the next decision has to do with hardware selection within the class (Table 17-2). The options available may be limited if the decision is to rely on a general-use facility, especially if the selection is limited to facilities within the investigator's own institution. However, even in such cases there is usually room for a choice if the institution has multiple general-use facilities. A comparative evaluation, including the use of benchmarking techniques to assess the computing power and cost of candidate facilities, is needed to make an informed choice. Consideration should be given to the experience of staff in the computing facilities in database management and data analysis and to the kinds of software packages available for those activities.

The existence of good database management packages, along with standard analysis packages, such as provided in BMDP, SPSS, and SAS (Devan and Brown, 1979; Dixon, 1981; Norusis, 1983; Ray, 1982), can markedly reduce

Table 17-1 General-use versus dedicated computing facilities

I. General-use facility

A. Pros and cons

- Likely to provide more computing power for the study than is feasible with a dedicated facility, but access to the facility may be limited
- Investigators are freed of responsibilities for operation of the facility; however, the operators of a general-use facility may be insensitive to specific needs of the trial
- Number of programming options on a general-use facility is likely to be greater than on a dedicated facility
- Generally provides a wider array of hardware than available on a dedicated facility
- Protection of data files on the system may be more difficult than with a dedicated facility

B. Factors favoring choice of general-use facility

- Existence of good general-use facility operated by staff responsive to user needs and equipped with hardware needed for the study
- Total duration of the trial, including the period of final analysis, relatively short (e.g., ≤ 3 years)
- Programming and data processing staff needed for the trial is small (e.g., ≤ 1 FTE)
- No one in the data center staff has the interest or talents needed for operation of a dedicated facility

II. Dedicated facility

A. Pros and cons

- Access to computer can be limited to study personnel, thereby avoiding competition with other users
- Limited access may make it easier to protect data files against unauthorized entry
- Amount of computing power and number of hardware and software options likely to be more limited than on large general-use facilities
- Responsibility for operation of the facility rests with study personnel. May be a disadvantage depending on the skills and interests of the personnel involved

B. Factors favoring choice of dedicated facility

- No general-use facility in the institution housing the data center, or the facilities that exist are overloaded
- Data processing needs are sizable and will continue over a long period of time (e.g., > 3 years)
- Programming and data processing staff needed for the trial is fairly large (e.g., ≥ 4 full-time equivalents)
- The existence of staff with the interest and talents needed for operation of a dedicated facility

Table 17-2 Considerations in choosing among computing facilities

A. Considerations in choosing among different general-use facilities

- Type and amount of staffing available for advice and consultation
- Hours of operation and modes of access (e.g., only on-site batch entry versus entry via remote job entry station or via CRT work station)
- Record of mainframe hardware supplier (e.g., firm with an established record for sales and service versus one that is a recent entry into the hardware field)
- Primary use of the facility (e.g., research versus administration)
- Compatibility of hardware and software features with other facilities (especially important if there is a need to switch facilities during the trial)
- Array of available hardware and software packages, particularly for data management and data analysis
- Past history of operation, including record of past hardware upgrades
- Level of satisfaction expressed by other research users of the facility
- Charging policy for computer time, on-line data storage, printing, etc.

B. Choosing among different dedicated facilities

- Available hardware and software features, especially those related to computing power, response time, database maintenance, and construction of files for data analysis
- Compatibility of programming languages with other operating systems
- Past history of vendor in producing and servicing small-scale dedicated computers
- Nature of details contained in manuals for operating the facility
- Vendor method of providing updates to the system and their costs
- Expertise of vendor sales and service personnel
- Level of access to vendor systems personnel for answering questions having to do with operation of the system
- Cost and maintenance charges

the amount of programming time required for both kinds of activities.

The options available if a dedicated facility is chosen are greater and more varied. Making an informed judgment may require months of work to collect the necessary cost and operating information. Highly specialized items of equipment, requiring use of esoteric programming languages, should be avoided. The cost and inconvenience involved in converting programs to operate on some other system may make it impractical to consider conversions later on.

A crucial cost issue is whether to purchase or lease the required hardware. Generally, purchase is cheaper than lease for items used at least three years. The disadvantage is that purchase may make it impractical to take advantage of subsequent upgrades, especially if the upgrades involve new product lines.

17.3 ORGANIZATION OF PROGRAMMING RESOURCES

The requirements for the data system should be developed by data processing personnel, in collaboration with the clinical investigators. Development of programs should not be started until there is general agreement on the requirements for data flow and editing. It may be efficient to vest responsibility for the development and maintenance of programs needed for operation of the database and those needed for data analysis with different groups (e.g., see Meinert et al., 1983). The majority of programming work early in the trial will be related to development of the data management system. The demand for this will diminish once the basic database management systems are in place. Programming efforts thereafter will be limited to those needed for maintenance of the system and for implementing changes dictated by hardware or software changes or by modifications to the study protocol. The demand for analysis programming will begin once recruitment is under way. The first efforts in this regard will relate to analyses needed for performance and safety monitoring and later on for manuscript preparation. The overall demand for programming is likely to increase over the course of the trial.

The time spent in improving the efficiency of operating programs should depend on the number of times they are likely to be used over the course of the study, the amount of time required to run them, and the way computer charges are billed. Most general-use computing centers have charges for on-line data storage, number of lines printed, tape or disk I/Os, etc. Minor changes in the charging algorithm can have major cost implications for the trial. Reprogramming may be necessary to lessen their impact.

A major issue in the development of any system has to do with the amount of testing that is done before programs are released for use in the trial. Many flaws can be detected via the reviews that are part of any good programming effort. On-line testing should not be started until there has been a successful "walk-through" of the program. A number of test runs should be made thereafter. The data sets used for this purpose should be typical of data likely to be collected as part of the trial. A number of different data sets should be used to reflect a variety of conditions.

Operating programs should be sufficiently well documented to allow someone unfamiliar with the programs to operate them. The need for good documentation, although greatest in long-term trials because of the changes in programming personnel that can occur, is important for all trials. Use of a structured programming language, such as PL/I, can help in this process; however, there is no substitute for the critical review of others in testing the adequacy of the documentation.

17.4 OPERATIONAL REQUIREMENTS FOR DATABASE MAINTENANCE

Data will be added to the analysis database in blocks. Keyed data are usually stored in a temporary file until a defined data entry session has been completed or until after the close of a defined time period. Thereafter, the resulting data block is transmitted to the analysis database for storage and subsequent manipulation. The update schedule will depend on the rate of data flow and on how and where data are keyed. The data center in the Coronary Artery Surgery Study (CASS) gathered information keyed and temporarily stored at the clinics, by polling clinic workstations (usually at night) on a weekly basis. The Coronary Drug Project (CDP) updated its main database about every two weeks (Meinert et al., 1983).

The prime function of the updating process is to link new data with that already in the analysis file. This may be accomplished by physically locating new data for a patient next to that already on file for the patient or by use of directories in which new data are added to the end of the file without regard to location of other data pertinent to a particular patient. The approach used will be determined by the type of computing hardware and software features available and the cost of data retrieval under one structure versus another.

The computer data file should be designed to minimize the amount of sorting and hand processing preparatory to an update, as well as the amount of computer time needed for the update. Generally, files that are constructed for easy updating are not easy to use for data analysis. Hence, it is usually necessary to reorganize them preparatory to any analyses.

A crucial issue in the updating process has to do with the disposition of data items that are still in a state of flux because of outstanding edit queries (see Chapter 16). Should such items be added to the analysis database or should they be excluded until the edit queries have been resolved? The CDP analysis database excluded all such data items. They were added to the file, on an item-by-item basis, as they cleared the edit process. They were included in the Aspirin Myocardial Infarction Study (AMIS) analysis database. However, items with outstanding edit queries were flagged. The flags remained in place until the edit queries were resolved and were used to eliminate questionable data for certain of the analyses performed.

17.5 DATA SECURITY PRECAUTIONS

The database of the study must be safeguarded against loss or unauthorized use (see the next section and Sections 15.5 and 24.4 for comments concerning storage of the original study records). Table 17-3 provides a list of the general precautions and safeguards that should be taken in any data operation (Part A), a list of safeguards applicable to files containing patient identifying information (Part B), general methods for protecting data files against misuse (Part C), and methods for protecting files against loss or destruction (Part D).

It is the responsibility of the study leadership to outline data security guidelines for the trial and to make certain that they are followed. Staff should be instructed as to their duties and responsibilities regarding data safeguards before they are allowed access to any study data. They should be cautioned against the release of data to anyone except authorized individuals, and then only through approved channels. All employees concerned with data processing should be given instructions regarding data security and should be informed (perhaps via statements they sign) of the types of disciplinary actions, including immediate dismissal, that can be expected if those safeguards are ignored or willfully violated.

Several of the large-scale multicenter trials (e.g., Aspirin Myocardial Infarction Study, Macular Photocoagulation Study, and Persantine Aspirin Reinfarction Study) have data systems that preclude collection of any personal identifiers, such as patient name and address, at the data center. The proscription provides a means of eliminating any chance for breaches of patient confidentiality in the data center (see Part B of Table 17-3 for safeguards used when patient identifying information is collected).

The data center has a responsibility to protect data in its custody against loss or destruction, whether caused by mistakes, accidents, or purposeful acts. A good data center will have the capability of regenerating the analysis database via backup files. Ideally, the tapes or disks containing these files should be stored in a building remote from the one housing the main database. At least two sets of backup tapes (or disks) should be maintained so that one set can be held in reserve while the other is used to restore the system. The schedule for generation of updated tapes or disks for backup purposes will be a function of the rate at which new information is added to the analysis database. (See Meinert et al., 1983 for a description of the CDP backup system.)

17.6 FILING AND STORING THE ORIGINAL STUDY RECORDS

The clinic should retain a copy of all data forms and related records generated in the trial until all essential work, including final analysis of the results, has been completed. This file may be the only hard copy of study records that exists. This will be the case in single-center trials without data centers and in multicenter trials with distributed data entry (see Section 16.2). Generally, a second paper file is needed if data entry is done outside the clinic, especially in multicenter trials. The file used for data entry should be considered the official file of the study and should contain the original copy of all paper forms and related records.

Table 17-3 Precautions and safeguards for database operations

A. General precautions and safeguards

- Study leadership that is sensitive to needs for data security
- Staff experienced in the operation of a database and in protecting it against loss or misuse
- Signed assurance from each employee authorized to work on the database, stating he understands the safeguards and precautions to be followed and the consequences of a willful disregard of them
- Periodic staff meetings to remind database personnel of required operating procedures and safeguards
- Periodic review of required operating procedures and established safeguards by study leaders
- Monitoring for adherence to precautions and safeguards via periodic on-site checks

B. Patient confidentiality safeguards

- Data flow procedures from the clinic to the data center that exclude transmission of patient identifying information
- Electronic storage of patient identifying information in enciphered form or in a separate file
- Separation of the file containing patient identifying information from other files
- Physical separation of pages containing personal identifying information from other pages of the data forms (especially if forms contain highly sensitive information)
- Proscription against distribution of data listings that contain patient name, name code, or any other identifiers easily associated with a specific patient
- Proscription against use of patient name, name code, hospital chart or record number, or other unique identifiers, such as Social Security number, in any published data listing. Study ID number should not be published if it is possible for people outside the study to use that number to identify a patient. Published UGDP patient listings (University Group Diabetes Program Research Group, 1970e, 1975, 1977, 1982) were devoid of both clinic and patient ID number for this reason.
- Secure procedures for disposing of computer output from aborted runs that contain patient identifying information
- Denial of access to any patient record stored in the data center to persons outside the data center without the express written consent of the patient

C. Safeguards against misuse

- Limit the number of persons in the data center who have access to the original study forms or any related data file, especially those containing patient identifying information
- Restrict access to the analysis computer files containing study results through use of passwords or other means
- Proscribe release of any data listing, tape, etc., without approval of the study leadership committee
- File completed study forms, data tapes, and disks, in an attended, locked area

D. Loss safeguards

- Maintain a duplicate file of the original study records (e.g., by requiring clinics to maintain a copy of forms and records sent to the data center)
- Microfilm original data forms, computer listings, study manuals, meeting minutes, etc., for storage in a secure location
- Establish and maintain a series of backup tapes (or disks) for the analysis database that will allow restoration of it in the event of a system malfunction
- Store copies of backup tapes (or disks) of the main analysis database in an off-site location or in an onsite fireproof vault
- Establish strict rules to safeguard access to backup tapes (or disks) to avoid unauthorized use in restoration efforts
- Provide backup tapes (or disks) of all essential programs, such as those needed for editing, inventorying, storage, retrieval, and analysis of the study data, as well as programs used for the operating systems
- Carry out occasional "fire drills" to test the ability of the staff in the data center to restore the main analysis database from backup tapes (or disks)

Decisions must be made as to where to house records that cannot be easily or reliably reproduced, such as X rays. Records that are needed for patient care should remain in the clinic or be returned to it as soon as they are read and the information from them has been codified and keyed. Some records, such as ECG tracings, can be "duplicated" by making a second tracing when the patient is examined. However, this option does not exist if the "duplication" entails added risks for the patient (e.g., as with X rays).

Both the official and backup paper files should be stored in locked cabinets in a secure area. The files should be checked periodically to make certain needed updates are made and that they do not become cluttered with superfluous materials.

The organization of the file will depend on where the file resides and how it is to be used. Those housed in the clinic will almost certainly be organized along patient lines. Those housed at the data center may be organized in other ways. For example, the CDP Coordinating Center found it convenient to arrange paper records by form type and by edit period (i.e., time period in which the forms were received). This ordering was more efficient than an arrangement by patient ID number and visit because of the data entry and editing process used by the center.

Data forms and related records stored at the clinic and data center may be retained in their original state or on microfilm. If microfilm is used, the original records should be retained until microfilm images have been checked for legibility and proper identification. Destruction of study forms and related records should be in accordance with local statutes for medical records. Data forms, medical records, computer listings, or microfilm images that contain patient identifying information should be burned or shredded. They should not be moved to the disposal site unless they can be destroyed upon receipt.

General National Institutes of Health guidelines require investigators to retain raw study documents (or microfilm copies of them) for a minimum of two to three years after expiration of funding (Department of Health, Education and Welfare, 1976; Department of Health and Human Services, 1981, 1982b). Requirements may extend beyond these limits in any case where there are legal challenges to the study, or where the results are under review by some official government agency. Prudent investigators will retain study records well beyond the required legal limit for scientific reasons alone.

17.7 PREPARATION OF ANALYSIS TAPES

Most data analyses will be done from a tape or disk created from the analysis database. There are several reasons for doing so, especially for interim analyses done for performance or safety monitoring (see Section 16.7 and Chapter 20). The principal ones are:

- To allow database maintenance personnel to continue making updates to the database without altering the analysis database
- To reduce the number of times the database is accessed for data analyses (in order to minimize the chances of programmer errors)
- To enable analysis personnel to rearrange data, including application of data reduction and special coding routines, in order to create a file that is more compact and suitably arranged for use with data analysis programs

Theoretically, the updating process could be terminated while data analyses are being done. However, termination of the updating process is not always practical, particularly when data analyses take weeks to carry out, as may be the case when preparing complex reports for patient safety monitoring (see Chapter 20). In any case, the interruption of data flow into the database complicates management of the updating process and reduces the usefulness of edits carried out in conjunction with the updating process.

It is wise to decide on a target date for generation of the analysis tape. The date chosen should correspond to the last major update or change to the analysis database or to some other event in the trial, such as close-out of follow-up or termination of a treatment. The format of the analysis tape or disk requires careful thought. Organization of data may be quite different from that of the analysis database. A decision must be made as to whether to array data by patient or by variable. Thought is also needed regarding the degree to which data are to be reduced as they are written onto the analysis tape or disk. Verbatim listings from the analysis database will provide the analyst with the greatest amount of flexibility, but they are also more complicated to use. Generally, some reduction, in which codes are combined to reduce the number of categories and by averaging aliquot determinations or repeat readings, will be necessary.

A decision is also needed regarding the amount of editing to be done on data written onto the analysis tape (or disk). Outlier values or values known to be in error should be identified when the tape is written to keep the analyst from having to perform these checks each time a variable is used.

18. Data analysis requirements and procedures

Another difficulty about statistics is the technical difficulty of calculation. Before you can even make a mistake in drawing your conclusion from the correlations established by your statistics you must ascertain the correlations.

George Bernard Shaw

18.1 BASIC ANALYSIS REQUIREMENTS

The essence of a trial emanates from comparisons of the treatment groups for differences in outcome. Those comparisons should be made following ground rules listed below.

Ground rule number 1 Patients used in treatment comparisons should be counted in the treatment group to which they were assigned.

Ground rule number 2 The denominator for a treatment should be all patients assigned to that treatment.

Ground rule number 3 All events should be counted in the comparison of primary interest.

Clearly, there are situations in which the first rule is followed, but the second is violated (e.g., certain patients are excluded from analyses because their treatment was not in "accordance" with the study protocol). The third rule is an admonition against analyses in which investigators elect to present results only for events believed to be related to the disease process under

Table 18-1 Examples of analysis ground rule violations

Violation	*Example*
Counting only a portion of the events observed	Carrying out the primary analysis for cause specific mortality, ignoring all cause mortality
Counting only those events that occur after a specified period of treatment	Restricting the database for the primary analysis to 30-day postsurgical deaths in a surgery trial, or by ignoring deaths that occur within a specified time period after the initiation of treatment in a drug trial
Using only those patients who received their assigned treatment or who had perfect (or suitably high) adherence to the assigned treatment	Exclusion of patients from the database who did not receive the "full" course of treatment in a drug trial
Allowing the treatment actually administered to determine the group in which a patient is counted	Counting a patient allocated to control treatment as a member of test-treated group because he received the test treatment
Using only "evaluable" patients	A cancer trial that ignores results for patients who failed to develop tumors of a certain size
Exclusion of ineligible patients enrolled in the trial	Elimination of patients who were judged ineligible after enrollment by personnel who were aware of treatment assignment and course of treatment

treatment (e.g., cardiovascular deaths in a heart study). See Table 18-1.

An unsophisticated investigator can be expected to rebel at the notion of using data from patients who refused the assigned treatment or who were not treated in accordance with the study protocol for making treatment comparisons. One temptation is to ignore such patients and to proceed with analyses as if they were never enrolled—a violation of the second ground rule. The only clue offered to readers to indicate that this was done may be a single telltale sentence, such as "The analyses in this paper have been restricted to evaluable patients." Of equal concern are cases where data from all patients are used, but where the primary analysis is done by the treatment administered rather than by the one assigned— a violation of the first ground rule. The main reason for randomizing in the first place, as noted in Chapter 8, has to do with the desirability of establishing treatment groups that are free of patient and physician selection bias. There is no assurance in this regard if patients are arbitrarily excluded from consideration after randomization.

Even if investigators accept the need for analyses based on the first two ground rules, they may willfully violate the third one. Counting rules that call for exclusion of certain events are, at best, difficult to defend because of their arbitrary nature. Further, their use can open the study to serious criticism. The Anturane Reinfarction Trial (ART) is a case in point. The published report from the trial drew criticism because of the failure of study investigators to count deaths occurring within 7 days of the initiation of treatment (Anturane Reinfarction Trial Research Group, 1978; Temple and Pledger, 1980). These exclusions made it difficult to interpret the mortality results.The concern of critics stemmed from uncertainty regarding the validity of the assumption underlying the exclusions (i.e., that deaths occurring in this time period were not treatment-related) and the apparent post hoc nature of the 7-day rule. Clearly, rules for exclusions devised after the start of data collection must be viewed with skepticism. The same is true for any exclusion rule, regardless of when it was written, which is administered by personnel who have access to patient treatment assignments, especially if subjective judgments are required in administering the rule.

Adherence to the above ground rules can lead to an underestimate of the true treatment effect, especially if treatment compliance is low, there are a lot of treatment crossovers (see Glossary for definition), or the denominators for the treatment groups include a lot of patients who could not be followed for the outcome of interest. The latter should not be a problem in trials using mortality as the outcome (see Chapter 15), but can be in trials with a nonfatal event or a labora-

tory or physiological measure as the outcome. A prudent investigator will carry out supplemental analyses aimed at quantifying the degree of conservatism implied. Certainly, there is no proscription against such analyses so long as they are accompanied by the primary ones suggested above. They may include analyses by level of treatment adherence and for a number of secondary outcomes as well.

18.2 BASIC ANALYTIC METHODS

This section provides a review of analytic methods used for making treatment comparisons in trials with a clinical event as the primary outcome. Readers may consult textbooks such as those by Armitage (1971), Brown and Hollander (1977), Bulpitt (1983), Buyse et al. (1984), Flandt-Johnson and Johnson (1980), Fleiss (1981), Ingelfinger et al. (1983), Kalbfleisch and Prentice (1980), Lee (1980), Pocock (1983), Shapiro and Louis (1983), and Tygstrup et al. (1982); and papers by Cutler and Ederer (1958), Kaplan and Meier (1958), Mantel and Haenszel (1959), Mantel (1966), and Peto et al. (1976, 1977), among others, for additional details.

18.2.1 Simple comparisons of proportions

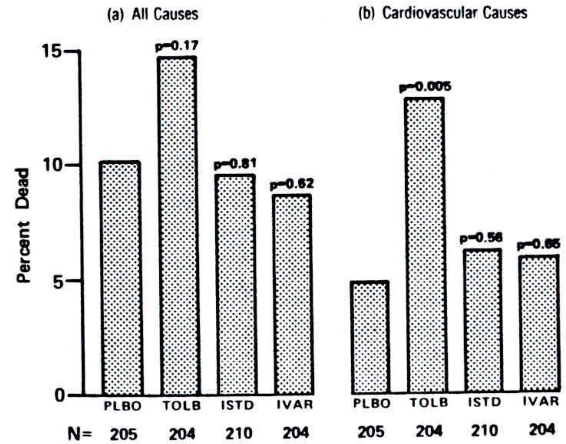
The simplest and often most useful analysis involves a comparison of the proportion of patients in the two treatment groups who have experienced the event of interest. This method of analysis is valid so long as:

- Patients in the treatment groups were enrolled over the same time period and are subject to the same intensity of follow-up
- The loss to follow-up is low and is the same across treatment groups
- The treatment groups have comparable baseline characteristics

Outcome analyses based on comparisons of proportions appear throughout publications of the trials sketched in Appendix B. Figure 18-1 is based on UGDP mortality data reported in a 1970 publication on tolbutamide (University Group Diabetes Program Research Group, 1970e).

This method of analysis, while best suited to binary data, need not be limited to such data if investigators are willing to convert a polychotomous or continuous outcome measure to binary form, as in the National Cooperative Gallstone Study (NCGS). Investigators in that study chose to categorize gallstone dissolution data as an all-or-none phenomenon for the primary analysis, even though the underlying measure was continuous (National Cooperative Gallstone Study Group, 1981a). Investigators in the Macular Photocoagulation Study (MPS) used a binary outcome (based on a comparison of base-

Figure 18-1 Number of deaths in the UGDP through October 7, 1969, by treatment group.



Note: p values recorded above the bars are based on $\chi^2_{(1\ df)}$ for the indicated drug-placebo comparison. The numbers of patients in the treatment groups are indicated below the bars.

Source: Reference citation 468. Adapted with permission of the American Diabetes Association, Inc., New York.

Table 18-2 Percentages of UGDP patients with indicated baseline characteristics (denominators given in parentheses)

Baseline characteristic	*PLBO*	*TOLB*	*p-value**
Age at entry ≥55	41.5(205)	48.0(204)	0.18
Male	30.7(205)	30.9(204)	0.97
Nonwhite	49.8(205)	47.1(204)	0.59
Fasting blood glucose ≥110 mg/100 ml	63.5(203)	72.1(204)	0.07
Relative body weight ≥1.25	52.7(205)	58.8(204)	0.21
Visual acuity (either eye ≤20/200)	4.3(188)	5.2(192)	0.66

Source: Reference citation 468. Adapted with permission of the American Diabetes Association, Inc., New York.
*Probability of chi-square value as large as or larger than the one observed under the null hypothesis

line and follow-up visual acuity readings) instead of mean change in visual acuity as the principal outcome measure (Macular Photocoagulation Study Group, 1982, 1983a, 1983b).

Furthermore, use of this mode of summary is not limited to outcome measures. It is useful in characterizing differences in the baseline composition of treatment groups and for comparisons of various kinds of follow-up data as well. Table 18-2 is an example of a comparison of the distribution of selected baseline variables that have been converted to binary form (University Group Diabetes Program Research Group, 1970e). Table 18-3 illustrates use of proportions in summarizing follow-up data on observed side effects (Persantine Aspirin Reinfarction Study Research Group, 1980b).

Statistical evaluation of the difference observed via a comparison of proportions can be performed using Fisher's exact test (Fisher, 1946; see also Chapter 9). The p-value for the test corresponds to the probability of obtaining a test-control difference as large or larger than the one observed under the null hypothesis of no difference. The p-value may be obtained using packaged computer programs for the test or from tables, such as those constructed by Lieberman and Owen (1961).

The continuity corrected chi-square approximation to the test can be used if the numerators for the two percentages being compared are both ≥ 5 and the denominators are ≥ 30. The p-values obtained in such cases are indistinguishable from those obtained with Fisher's exact test. In fact, the approximation is reasonably good even if denominators are as small as 20 (Cochran, 1954).

Table 18-3 Percentages of PARIS patients with indicated complaint during follow-up

Complaint	*Treatment group*		*Z-value*
	PR/A	*PLBO*	
Stomach pain	15.8	7.7	3.74
Heartburn	9.6	5.2	2.58
Vomiting	2.5	1.0	1.59
Denominator	810	406	

Source: Reference citation 376. Adapted with permission of the American Heart Association, Inc., Dallas, Texas.

18.2.2 Lifetable analyses

The typical trial involves patient recruitment over an extended period of time and follow-up through a common calendar time point. Hence, any analysis done during or at the end of the trial will involve patients with varying lengths of follow-up, depending on when they were enrolled. Simple counts of events, such as shown in Figure 18-1, are not designed to take account of follow-up time and hence are insenitive to the way events accumulate over time. The cumulative proportion of patients experiencing events can be the same even though there are marked differences between the treatment groups as to when events occur over the course of follow-up, as illustrated in Table 18-4 for a hypothetical trial. Note that comparisons of the percent dead based on tabulations done at the end of calendar year 6 or before favor treatment B. Those done at the end of calendar year 7 and thereafter favor treatment A.

One way of tracking changes over time via proportions is illustrated in Figure 18-2. This method of analysis, while useful for safety monitoring (see Chapter 20), does not give a means of characterizing the treatment groups with regard to the rate of occurrence of events. Rate calculations are ordinarily made using lifetable methods (such as described by Elandt-Johnson and Johnson, 1980; Kalbfleisch and Prentice, 1980; Lee, 1980), as illustrated in Figures 18-3 for the UGDP and 18-4 for the CDP. Other examples may be found in publications from the Aspirin Myocardial Infarction Study (AMIS), Hypertension Detection and Follow-Up Program (HDFP), Multiple Risk Factor Intervention Trial (MRFIT), and PARIS (see Appendix B for references).

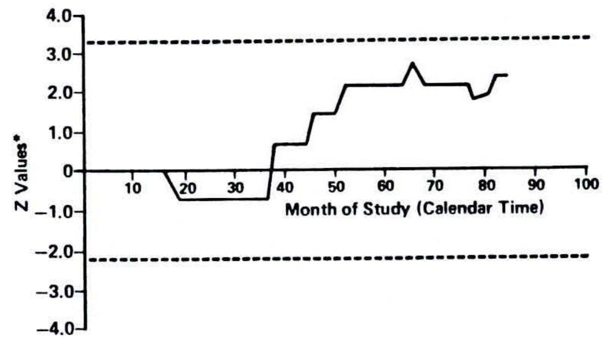
The main advantage of the lifetable approach is that it provides a means of dealing with varying lengths of follow-up, as illustrated in Table 18-5. The cut-off date for the analysis was October 7, 1969. All patients by that time had been under follow-up for a minimum of 3 years, 8 months and a maximum of 8 years, 8 months. Hence, the only attrition during the first 3 years of follow-up was that due to death. Thereafter, it was due to both deaths and withdrawals because of when patients were enrolled. For example, there were five patients in the tolbutamide-treated group who were enrolled after October 7, 1965, and who were still alive on October 7, 1969. They were counted as withdrawals during the fourth year of follow-up since they had not

Table 18-4 Hypothetical trial involving comparison of percentage of patients dead at indicated time points*

Calendar time from start of trial	*Cumulative number of patients enrolled*		*Cumulative percent dead*	
	Treatment A	*Treatment B*	*Treatment A*	*Treatment B*
1 year	100	100	10.0	2.0
2 years	200	200	13.6	3.5
3 years	300	300	16.2	6.5
4 years	300	300	21.0	12.3
5 years	300	300	24.2	19.4
6 years	300	300	26.7	25.8
7 years	300	300	28.9	31.7
8 years	300	300	31.1	37.2
9 years	300	300	33.1	42.2

*Percentages calculated assuming annual mortality rates (per 100 population) of 10, 8, 5, 4, 3, 3, 3, 3, and 3 for years 1 through 9 of follow-up, respectively, for treatment group A and 2, 3, 8, 8, 8, 8, 8, 8, and 8 for treatment B. Enrollment is assumed to have taken place on the first day of years 1, 2, and 3.

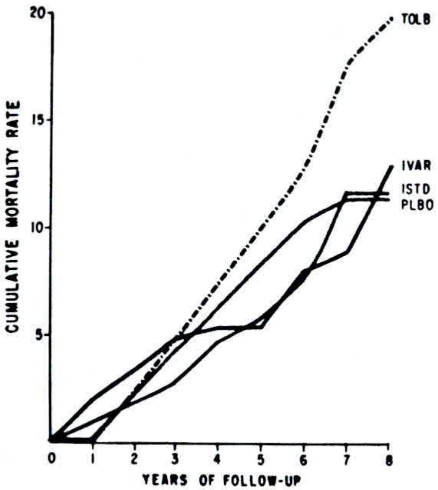
Figure 18-2 Plot of observed ESG1-placebo difference in percent of CDP patients dead from lung cancer.



*Z values plotted are for observed ESG1-placebo differences in proportions of deaths from lung cancer. Dotted lines denote Z values corresponding to 0.05 level of significance taking into consideration there were repeated evaluations of the data for treatment differences over the course of the trial.

Source: Reference citation 105. Adapted with permission of the American Medical Association, Chicago, Ill. (copyright © 1973).

Figure 18-3 UGDP cumulative lifetable mortality rates by year of follow-up and by treatment assignment.

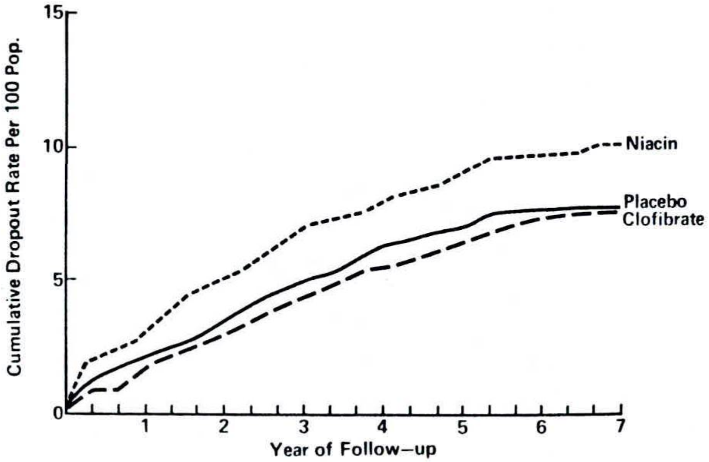


Source: Reference citation 468. Reproduced with permission of the American Diabetes Association, Inc., New York.

been in the study long enough to have completed the fourth year of follow-up.

Statistical comparisons of lifetable rates may be done using confidence estimation or log rank tests. The plot of lifetable rates reproduced in Figure 18-5 uses two standard error limits (i.e., approximate 95% confidence intervals) about the line of no difference to assess the statistical importance of the DT4-placebo mortality difference. The log rank test summarized in Table 18-6 is for data given in Table 18-5. (See Mantel and Haenszel, 1959, Mantel, 1966, and Peto et al., 1977 for general details regarding the test.) Ideally, the calculations should be based on exact time to death, rather than on grouped data, as given in Table 18-5. However, the difference between the two methods of calculation will be small provided the deaths are uniformly distributed within the intervals and that they are not concentrated in just one or two of the intervals. The difference in this example is trivial. Use of exact time to death yielded a log rank test value of 1.82 as contrasted with a value 1.78 for grouped data.

Figure 18-4 Lifetable cumulative dropout rates for the clofibrate, niacin, and placebo treatments in the CDP.



N = 4,706 4,421 4,167 3,911 3,651 2,336 729

Note: N denotes total number of patients in clofibrate, niacin, and placebo groups combined. Approximate numbers for individual treatment groups are 2/9, 2/9, and 5/9 times N for clofibrate, niacin, and placebo, respectively.

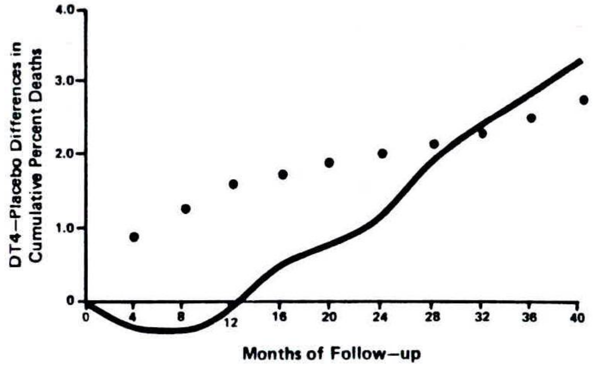
Source: Reference citation 107. Adapted with permission of the American Medical Association, Chicago, Ill. (copyright © 1975).

Table 18-5 Lifetable cumulative mortality rates for the placebo and tolbutamide treatments in the UGDP, as of October 7, 1969

Year of follow-up	*Number of deaths in*		*Number of survivors in*		*Total number starting interval*	*Estimated probability of death in interval*	*Observed rate per 100 population at risk*		*Mortality rate standard error*
	Patients due for withdrawal in interval	*Patients not due for withdrawal in interval*	*Patients due for withdrawal in interval*	*Patients not due for withdrawal in interval*			*Mortality rate*	*Survival rate*	
Placebo treatment group									
First	0	0	0	205	205	0.0	0.0	100.0	0.6
Second	0	5	0	200	205	0.024	2.4	97.6	1.1
Third	0	4	0	196	200	0.020	4.4	95.6	1.4
Fourth	0	4	4	188	196	0.021	6.4	93.6	1.7
Fifth	0	4	23	161	188	0.023	8.5	91.5	1.8
Sixth	0	3	43	115	161	0.022	10.4	89.6	2.1
Seventh	0	1	50	64	115	0.011	11.4	88.6	2.5
Eighth	0	0	36	28	64	0.0	11.4	88.6	2.8
Tolbutamide treatment group									
First	0	0	0	204	204	0.0	0.0	100.0	0.6
Second	0	5	0	199	204	0.025	2.4	97.6	1.1
Third	0	5	0	194	199	0.025	4.9	95.1	1.4
Fourth	0	5	5	184	194	0.026	7.4	92.6	1.7
Fifth	0	5	24	155	184	0.029	10.1	89.9	1.8
Sixth	1	3	41	110	155	0.030	12.8	87.2	2.1
Seventh	0	5	47	58	110	0.058	17.8	82.2	2.5
Eighth	0	1	33	24	58	0.024	19.8	80.2	2.9

Source: Reference citation 468. Adapted with permission of the American Diabetes Association, Inc., New York.

Figure 18-5 CDP lifetable plot of the DT4-placebo mortality differences and 2.0 standard error limits for the differences.



Source: Reference citation 103. Adapted with permission of the American Medical Association, Chicago, Ill (copyright © 1972).

18.2.3 Other descriptive methods

Any comparison of outcome by treatment group should be accompanied by other analyses to help in interpretation of the results of the trial: Tables 18-2 and 18-7 and Figure 18-6 provide examples of supporting analyses, as taken from the UGDP (University Group Diabetes Program Research Group, 1970e). The results in Table 18-2 are useful for assessing the baseline comparability of the treatment groups. Table 18-7 was used to characterize differences among treatment groups with regard to treatment adherence. Figure 18-6 provides a plot of changes in fasting blood glucose levels for the cohort of patients followed through 4.75 years (i.e., through 19 follow-up examinations). Only patients who remained under active follow-up over this time period were included in the analysis. A plot of means, based on the number of patients observed at each follow-up examination, might have been used instead. However, the two forms of analyses are not necessarily interchangeable. They will yield different results if the composition of the study population enrolled, with regard to the variable of interest, changed over the course of patient recruitment.

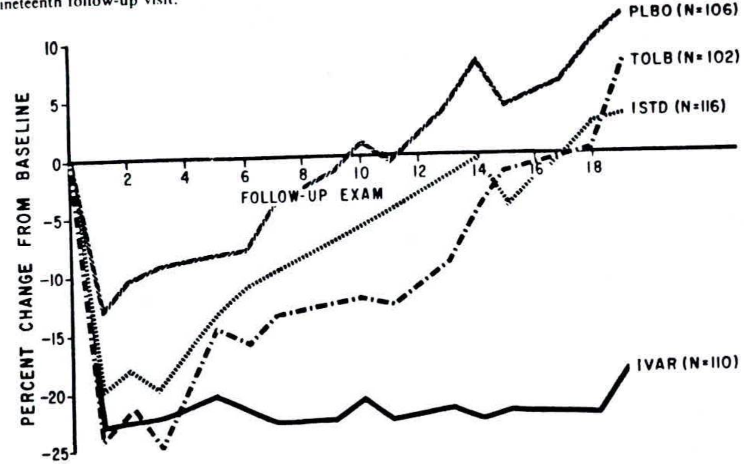
Table 18-6 Log rank test for comparing lifetables in Table 18-5

Year of Follow-up	*Number starting interval*			*Observed deaths*			*Expected deaths*		
	PLBO	*TOLB*	*Total*	*PLBO*	*TOLB*	*Total*	*PLBO*	*TOLB*	*Total*
0-1	205.0	204.0	409.0	0	0	0	0.00	0.00	0.00
1-2	205.0	204.0	409.0	5	5	10	5.01	4.99	10.00
2-3	200.0	199.0	399.0	4	5	9	4.51	4.49	9.00
3-4	194.0	191.5	385.5	4	5	9	4.53	4.47	9.00
4-5	176.5	172.0	348.5	4	5	9	4.56	4.44	9.00
5-6	139.5	134.5	274.0	3	4	7	3.56	3.44	7.00
6-7	90.0	86.5	176.5	1	5	6	3.06	2.94	6.00
7-8	46.0	41.5	87.5	0	1	1	0.53	0.47	1.00
Total				21	30	51	25.76	25.24	51.00

Source: Reference citation 441.

Log rank $\chi_1^2 = (21 - 25.76)^2/25.76 + (30 - 25.24)^2/25.24 = 1.78$, p-value $= 0.18$

Figure 18-6 Percent change in fasting blood glucose levels for cohorts of patients followed through the nineteenth follow-up visit.



Source: Reference citation 468. Reproduced with permission of the American Diabetes Association, Inc., New York.

18.3 ADJUSTMENT PROCEDURES

To be valid, the evaluation of treatment effects must be performed on treatment groups that are comparable with regard to baseline characteristics. Usually, the comparability provided by randomization is adequate. However, randomization does not guarantee comparability. As noted in Chapter 10, stratification can be used to assure comparability for a few variables, but the distribution with regard to others must be left to chance. As a result, there can be minor, and sometimes even major, differences in the baseline composition of the study groups. The impact of such differences on treatment comparisons should be removed using procedures such as those outlined below.

Table 18-7 Percentage distribution of UGDP patients by level of treatment adherence

*Level of adherence**	*Treatment group*			
	PLBO	*TOLB*	*ISTD*	*IVAR*
Low	10.2	10.3	12.8	14.8
Intermediate	20.0	15.7	29.6	39.9
High	69.8	74.0	57.6	45.3
Number of patients	205	204	210	204

Source: Reference citation 468. Reproduced with permission of the American Diabetes Association, Inc., New York.

*Defined as follows:

Low Patient took all of prescribed study medication <25% of all follow-up periods

Intermediate Patient took all of prescribed study medication 25–74% of all follow-up periods

High Patient took all of prescribed study medication ≥75% of all follow-up periods

18.3.1 Subgrouping

The simplest approach involves making the required treatment comparisons in subgroups of patients that are homogeneous for selected entry characteristics. This method of adjustment is illustrated in Table 18-8. All of the subgroups were formed using measures observed before the start of treatment. The table indicates the size of each subgroup and the percentage of patients in the subgroup who had died as of the analysis cut-off date, October 7, 1969.

This approach, while simple, has obvious limitations. Thirty-two (i.e., 2^5) different subgroups would be required to simultaneously categorize patients for the presence or absence of the five measures represented in Table 18-8. The number

Table 18-8 Percentages of patients dead within specified subgroups created using selected baseline characteristics

	Number		Percent dead	
Entry risk factor	*PLBO*	*TOLB*	*PLBO*	*TOLB*
Definite hypertension				
Absent	127	139	11.0	12.9
Present	74	60	9.5	16.7
History of digitalis use				
No	193	183	8.3	13.1
Yes	9	15	55.6	33.3
History of angina pectoris				
No	192	187	9.4	13.9
Yes	10	14	30.0	21.4
Significant ECG abnormality				
Absent	193	193	9.3	13.0
Present	6	8	33.3	50.0
Cholesterol				
<300 mg/100ml	181	169	10.5	14.8
≥300 mg/100ml	17	30	11.8	13.3
Any of above cardiovascular risk factors				
None	98	100	9.2	11.0
One or more	88	92	12.5	17.4

Source: Reference citation 468. Adapted with permission of the American Diabetes Association, Inc., New York.

of patients in many of the subgroups would be too small for meaningful comparison.

In addition, the method requires use of arbitrary cut-points for subgroupings involving continuous variables. The arbitrary nature of the cut-points selected can raise questions concerning the validity of the analyses presented, especially if there is any suspicion that they were chosen to minimize or maximize observed treatment differences.

18.3.2 Multiple regression

An alternative approach that avoids some of these problems and provides a means of controlling for several sources of variation simultaneously involves use of regression models represented by Equations 18.1 and 18.2. (See Cox, 1958, Draper and Smith, 1966, and Kleinbaum et al., 1982, for details on methods of estimation using the models.) The models are used to estimate the probability that a patient experiences the outcome of interest, given a particular set of entry characteristics. One drawback to the linear regression model has to do with the possibility of obtaining probability estimates that lie outside the range of 0 to 1. This possibility is avoided with the logistic model.

Linear[1] multiple regression model

$$y_i = A + \epsilon_i \tag{18.1}$$

Logistic multiple regression model

$$y_i = \frac{1}{1 + e^{-A}} + \epsilon_i \tag{18.2}$$

where

$A = \beta_0 + \beta_1 x_{1i} + \cdots + \beta_j x_{ji} + \cdots + \beta_k x_{ki}$

y_i = outcome observed for the ith patient (either 0 or 1 for binary outcome measures)

x_{ji} = value observed for the ith patient and jth entry characteristic ($j = 1, \cdots, k$)

ϵ_i = error associated with y_i

and

$\beta_0, \beta_1, \ldots, \beta_k$ = regression coefficients (parameters) to be estimated from observed data

The UGDP used a logistic regression model to adjust observed mortality results for differences in the distribution of 14 different entry characteristics (University Group Diabetes Program Research Group, 1970e). Results are summarized in Table 18-9. The CDP used both multiple linear and multiple logistic regression models to adjust observed mortality for as many as 54 different baseline characteristics (Coronary Drug Project Research Group, 1974, 1975).

The use of regression procedures for adjustment has been extended to event rates calculated from lifetables (Cox, 1972). The method has been used in studies such as AMIS (Aspirin Myocardial Reinfarction Study Research Group, 1980b) and PARIS (Persantine Aspirin Reinfarction Study Research Group, 1980b).

1. Referred to as linear because the model does not involve any parameter raised to a power other than unity. The term is not a comment on the shape of the curve arising from the analysis. The model may yield a curved line or surface depending on the form taken by the independent variable(s) in the model.

Table 18-9 Observed and adjusted tolbutamide-placebo difference in percent of patients dead

	TOLB	*PLBO*	*TOLB-PLBO difference*
Observed percent dead	14.7	10.2	4.5
Adjusted* percent dead	14.5	10.2	4.3

Source: Reference citation 468. Adapted with permission of the American Diabetes Association, Inc., New York.

*Based on logistic regression model using 14 different baseline characteristics.

18.4 COMMENT ON SIGNIFICANCE ESTIMATION

The p-values resulting from conventional tests of significance are often used by investigators to decide whether to characterize a particular result as being statistically significant. Clearly, p-values can help in the statistical quantification of a result, but they should not become a substitute for rational thought. The acceptance or rejection of a treatment rarely hinges on whether a difference reaches some arbitrary level of significance. In fact, the amount of evidence required to conclude that a test treatment is no better than the control treatment may be less than that required to conclude that it is better. Generally, there is need in the latter case to make certain the beneficial effects observed persist—a judgment that can be reached only by continuing follow-up for some time after the emergence of an important difference.

The question of what constitutes statistical significance is complex. Methodological problems involved in the interpretation of conventional tests of significance for safety monitoring are outlined in the next chapter. However, even if those problems are ignored, it is still necessary to use a good deal of caution in the interpretation of p-values. Most trials, even if designed to focus on a single outcome, will provide data on a variety of other outcome measures as well. For example, the CDP provided data on the rate of occurrence of myocardial infarctions, strokes, and several other nonfatal cardiovascular events, in addition to death. The p-values obtained for one outcome measure will not be independent of those obtained using another outcome measure.

19. Questions concerning the design, analysis, and interpretation of clinical trials

There are three kinds of lies: lies, damned lies and statistics.

Benjamin Disraeli

19.1 INTRODUCTION

This chapter focuses on questions concerning the design, analysis, and interpretation of study data. Material is presented in the form of questions and answers and is organized in categories related to the various aspects of a clinical trial.

19.2 QUESTIONS CONCERNING THE STUDY DESIGN

1a. *Question:* Can a new study treatment be added during the course of the trial?

Answer: Yes, but not without impact on the study design. The University Group Diabetes Program (UGDP) elected to add a fifth treatment, phenformin, 18 months after the start of patient enrollment (University Group Diabetes Program Research Group, 1970d). The allocation ratio of phenformin to tolbutamide to insulin standard to insulin variable to placebo was fixed at 3:1:1:1:1 and was satisfied after enrollment of every 14, 28, 42, etc., patient in each of the 6 clinics administering phenformin. Patients in the other 6 UGDP clinics and the first 32 patients in one of the clinics included in the phenformin portion of the study were allocated using a ratio of 0:1:1:1:1 in blocks of 16.

The two different allocation schemes created problems when treatment comparisons were made involving phenformin-treated patients (University Group Diabetes Program Research Group, 1975). The decision in the Coronary Drug Project (CDP) to study aspirin late in the trial avoided these design problems by setting up a separate trial using patients from discontinued treatments (Coronary Drug Project Research Group, 1976).

1b. *Question:* Can a treatment be deleted from the study design once the trial has started?

Answer: Yes. Use of the test treatment will have to be stopped if it is shown to be inferior to the control treatment. The control treatment will have to be stopped if it is inferior to the test treatment. The UGDP provides examples of the former kind of change (University Group Diabetes Program Research Group, 1970e, 1975). The Diabetic Retinopathy Study (DRS) provides an example of the latter type of change (Diabetic Retinopathy Study Research Group, 1976, 1978).

A treatment may also be deleted for reasons unrelated to treatment results. The original design of the DRS included a test treatment involving photocoagulation with both xenon arc and argon laser. The treatment was abandoned early in the course of the trial for practical reasons.

2a. *Question:* Do all clinics participating in a multicenter trial have to be in the trial from the outset?

Answer: No. Results from clinics can be combined regardless of when they were added to the trial, provided all clinics followed the same treatment protocol and all treatment assignments were made using a common allocation ratio. See question 1a.

2b. *Question:* What if a clinic in a multicenter trial resigns after it has started patient enrollment? Will the resignation affect treatment comparisons?

Answer: Clinic resignations are not uncommon. There were two in both the CDP and the National Cooperative Gallstone Study (NCGS) (Coronary Drug Project Research Group, 1973a; National Cooperative Gallstone Study Group, 1981a). They may be initiated by the clinic because of the death, illness, or departure of a key person or by the study leadership because of performance problems.

The loss of a clinic will reduce the overall precision of the trial unless other clinics are recruited to make up for the loss. The loss will be minimal if few patients are involved and if responsibility for the continued care and surveillance of patients already enrolled can be assumed by another clinic in the trial. It will be sizable if the clinic had a large number of patients that cannot be transferred to other clinics in the trial. Such patients will have to be counted as dropouts and treated as such for data analyses in the trial. A large number of dropouts caused by clinic resignations will make it difficult to detect treatment effects, but they should not invalidate treatment comparisons provided the allocation ratio in clinics that have resigned was the same as in the remaining active clinics. Incidentally, the possibility of clinic resignation in a multicenter trial is one reason why it is wise to construct the allocation schedule with clinic as a stratification variable.

3. *Question:* Is it proper to make modifications to the treatment protocol during the trial?

Answer: Many times it is not so much a question of propriety as of necessity. Changes must be made if patient safety is in question. Other changes may be necessary simply to clear up ambiguities in the protocol. All changes should be noted and reported in publications from the trial.

4a. *Question:* If the required sample size cannot be achieved, should it be reduced to bring it in line with reality?

Answer: It is always possible to find some combination of α, β, and Δ which yields the "desired" result (see Chapter 9). Reduction of the sample size via such manipulations, simply to bring it in line with expectation, is game playing.

4b. *Question:* How about revising the sample size calculation during the trial?

Answer: Revised sample size calculations, based on observed outcome and dropout rates, can help the investigators and sponsor decide if more clinics are needed or if the period of follow-up should be extended to achieve the desired statistical precision. The calculations should be made using the α, β, and Δ specified when the trial was planned (see Chapter 9).

4c. *Question:* Is it all right to change the outcome measure after the start of the trial as a means of reducing the sample size requirement?

Answer: Such maneuvers are open to the same criticism as mentioned in the answer to question 4a. One kind of maneuver involves a switch from a single event as the prime outcome measure to a composite event (see Glossary). The expected rate of occurrence of such an event will be higher than that for any of its component parts. The higher the expected rate, the easier it will be to detect a specified relative difference with a given sample size. However, the "gain" in precision is achieved at the expense of clinical relevancy. It is more difficult to interpret the meaning of a finding based on combinations of events than one that is based on a single set of events.

5. *Question:* Is it permissible to extend the period of patient follow-up to compensate for a lower than expected event rate in the control-treated group or for a shortfall in patient recruitment?

Answer: Yes.

6. *Question:* Is it necessary to specify stopping rules for the trial before it is started?

Answer: No. In fact, many trials are done without any formal stopping rules for reasons discussed in Chapter 20.

Other related questions: 7, 42, 43, and 47.

19.3 QUESTIONS CONCERNING THE SOURCE OF STUDY PATIENTS

7. *Question:* Is it all right to change patient eligibility criteria once the trial has started?

Answer: Ideally no, but some changes may be necessary. The likelihood of change is greatest in trials involving long periods of recruitment and in those in which investigators are having trouble meeting their sample size goals within the stated time periods. The changes will not affect the validity of treatment comparisons if they are independent of the observed treatment results and if the proportion of patients allocated to the different treatment groups remains unchanged over the course of patient enrollment.

8. *Question:* Will changes in the composition of the study population enrolled have an impact on treatment comparisons?

Answer: No, assuming the proportion of patients assigned to a particular treatment, relative to the total number of allocations made, remains constant over the course of the trial. This is usually assured with randomization procedures designed to balance the number of assignments made to the treatment groups at various points over the course of patient recruitment.

9. *Question:* Is it useful to collect data on patients screened for enrollment?

Answer: It is if there is a reliable way to define the base population at risk of enrollment, as in the Coronary Artery Surgery Study (CASS). The only patients considered for enrollment were those who had had a heart catheterization at a study clinic (Coronary Artery Surgery Study Research Group, 1981). It is not useful when the base population is ill defined, as in the UGDP. Investigators in that trial tried to maintain screening logs, but abandoned the effort because of lack of agreement among them as to who should be listed in the logs.

Other related questions: 73.

19.4 QUESTIONS CONCERNING RANDOMIZATION

10. *Question:* Is randomization needed for a valid trial?

Answer: Not necessarily, provided the method of assignment is free of treatment-related selection biases. In fact, some people have even argued that randomization is unnecessary (Harville, 1975; Lindley, 1982). Indeed, it would be if all extraneous sources of variation could be identified before the start of the trial and then controlled in the assignment process. However, this is rarely, if ever, possible. The main virtue of randomization is the protection it provides against patient or physician selection biases in the treatment assignment process.

11a. *Question:* Is it acceptable to use an informal, nonauditable method of random assignment, such as a coin flip?

Answer: Not if it can be avoided. Such methods, even if properly administered, are difficult to defend if questions are raised concerning the assignment process. There is no satisfactory way to dispel doubts concerning the possibility of selection bias with any nonauditable allocation scheme.

11b. *Question:* How about methods of randomization that base treatment assignment on a specified digit of the patient's Social Security or medical record number? Are they acceptable?

Answer: Again, not if they can be avoided. Most of these methods fail to satisfy the conditions needed for a sound allocation scheme, as discussed in Chapters 8 and 10.

12. *Question:* Are schemes such as those based on day of the week, time of day, or order in which patients are seen all right to use?

Answer: No. All such methods are susceptable to selection biases and, as a result, may not provide a valid basis for comparisons in the trial. It is too easy for patients or clinic staff to discover the assignment rules and then to alter the time or order in which patients are seen simply to achieve the "desired" assignments.

13. *Question:* Should the treatment assignment be blocked?

Answer: Yes. There can be subtle changes in the composition of the study population as the trial proceeds. Blocking helps to eliminate the impact secular changes may have on treatment comparisons (see Chapter 10).

14a. *Question:* Are the number adaptive schemes, such as the biased-coin method of randomization in which assignment probabilities change as a function of previous assignments, a substitute for blocking?

Answer: Yes. They can serve the same function, as suggested in Section 10.2.

14b. *Question:* Are such schemes better than those that rely on blocking to achieve the desired allocation ratio?

Answer: Yes and no. On the one hand, such methods avoid the problem of predictability as discussed in Chapter 10—a serious problem with small blocks of uniform size, especially in unmasked trials. On the other hand, they can yield longer unbroken runs of patients who are all assigned to the same treatment. Further, the schemes are more complicated to administer than schemes involving blocking.

15. *Question:* Should one use blocks of variable size if blocking is used?

Answer: Generally, yes, particularly in unmasked trials. The variation reduces the likelihood that clinic personnel will be able to predict a treatment assignment.

16. *Question:* Is it necessary to stratify on all important baseline variables in the randomization process?

Answer: No. Valid treatment comparisons can be made without any stratification.

17. *Question:* Is there a limit to the number of variables that can be controlled via stratification during the randomization process?

Answer: Definitely. Generally, it is not practical to stratify on more than two or three variables.

18. *Question:* Should one use clinic as a stratification variable in multicenter trials?

Answer: Generally yes, except in a situation in which there are so few patients per clinic (as in some multicenter trials involving an extremely rare disease) that it is impractical to do so. The characteristics of patients enrolled can vary widely from clinic to clinic. These differences, if uncontrolled, can confound treatment comparisons.

19. *Question:* Is there a way to determine whether randomization has "worked"?

Answer: No. A random process is defined by the methods underlying the process. The demographic and baseline characteristics of patients enrolled in the various treatment groups can be compared. However, the existence of a large difference involving an arbitrarily small p-value does not necessarily mean that the assignments were "nonrandom," nor that there was a breakdown in the way in which they were issued. The difference may be due to chance.

20. *Question:* Does the lack of baseline comparability among the treatment groups indicate a breakdown in the randomization process?

Answer: Not necessarily. It may be due to chance, as noted in question 19.

21a. *Question:* Is it all right for the data center to take back a treatment assignment once it has been revealed to the clinic?

Answer: No. The assignment and the patient for whom it was intended should be counted in the study once it has been disclosed. Care should be taken to make certain that the patient is eligible and willing to participate in the trial before the assignment is revealed (see Section 10.7).

21b. *Question:* Should returned assignments (assuming the envelopes in which they are contained have not been opened) be reissued?

Answer: They can be, but often are not because of the difficulties involved in reissuing them.

21c. *Question:* Can the returned assignments result in measurable departures from the desired allocation ratio?

Answer: Not if the number returned is small. They could if the number is large, but even in this case the chance of a sizable departure is small, unless the number is differential by treatment group—not likely except in cases where decisions to return assignments are made by personnel who know the treatment assignments when the decisions are made.

21d. *Question:* What if a mistake is made in preparing the assignment and the wrong one is disclosed to clinic personnel? Should it be taken back?

Answer: No. The assignment should stand as issued once it is disclosed.

21e. *Question:* Can such mistakes lead to a departure from the desired allocation ratio?

Answer: They should not, provided they are independent of treatment assignment. However, they can raise doubts regarding the integrity of the study if they occur frequently.

22a. *Question:* What if the clinic wants to return an assignment because it was used by mistake?

Answer: The assignment should stand as issued once it has been disclosed to clinic personnel.

22b. *Question:* What if a clinic wishes to switch a treatment assignment?
Answer: The assignment should stand as issued once it has been revealed to clinic personnel.

23a. *Question:* What if a clinic administers the wrong treatment to a patient. Should the assignment be changed to correspond to the treatment used?
Answer: No. The assignment should stand as issued. The mistake should be noted when the results of the trial are published.

23b. *Question:* Will mistakes of the type referred to in Question 23a affect the validity of the trial?
Answer: They may, depending on their frequency and whether they are treatment related.

24. *Question:* What if the observed allocation ratio departs from the one specified in the study design?
Answer: Small departures are to be expected, even with small block sizes, few allocation strata, and no returned assignments. Bigger departures can occur with large blocks and multiple strata. Generally, other than detracting from the esthetic quality of the allocation design, the departures will not affect the validity of the trial. An obvious exception is where the departures are treatment related.

25. *Question:* Is it a good idea to have a large number of allocation strata?
Answer: Yes and no. On the one hand, the greater the number of strata the greater the control of extraneous sources of variation. On the other hand, numerous strata will complicate management of the allocation process (see Section 10.3.2).

Other related questions: 7, 8, 49, 50, 51, and 56.

19.5 QUESTIONS CONCERNING MASKING

26. *Question:* Is an unmasked trial valid?
Answer: Masking per se is not an indicator of validity. Valid treatment comparisons can be made without masking. The issue is whether the data collection process, especially as it relates to outcome assessment, is subject to treatment-related biases.

27. *Question:* What if it is impossible to mask?
Answer: This is often the case. The trial should be designed recognizing the opportunities for treatment-related bias. Bias control procedures, such as those discussed in Chapter 8, should be considered.

28. *Question:* Are there circumstances in masked drug trials in which the treatment assignment for a specific patient must be revealed during the course of the trial?
Answer: Yes, a few. However, as noted in Section 8.5, they should be limited to emergency situations. The preferred approach is to terminate use of the assigned treatment without revealing its identity.

29. *Question:* Are there cases in which an entire set of assignments must be unmasked during the trial?
Answer: Yes, when a treatment is discontinued during the study. Clinic personnel will need to identify patients affected by the change in order to implement it.

30. *Question:* Should a patient be informed of the treatment assignment if he is separated from the trial before it is over?
Answer: The answer depends on when the separation occurs, on the arrangements agreed upon when the patient was enrolled, and on the health care needs of the patient. Unmasking individual patients as they depart from the study can create problems in maintaining the mask for other patients, as discussed in Section 15.4.

31. *Question:* Should patients in a masked trial be told of the treatments they were on when the trial is terminated?
Answer: Yes.

32. *Question:* Should the effectiveness of the treatment masking be assessed when the trial is over?
Answer: Yes, as discussed in Section 15.4. Guesses made by clinic staff and patients regarding treatment assignments can be used to make the assessments.

Other related questions: 40, 62, 63, and 64.

19.6 QUESTIONS CONCERNING THE COMPARABILITY OF THE TREATMENT GROUPS

33. *Question:* Are tests of significance helpful in identifying differences in the baseline characteristics of the treatment groups?
Answer: Yes, but the results of such tests must be viewed with caution because of the problems associated with making multiple comparisons, as mentioned in Section 9.3.12.

34. *Question:* When assessing treatment effects, is there a need to be concerned with differences in the baseline comparability of the treatment groups if the differences are small?
Answer: Probably not, but as noted in Section 18.3, it is a good idea to adjust for baseline differences even if small.

35a. *Question:* Is it reasonable to expect the treatment groups to have identical baseline distributions?
Answer: No. The groups will be identical only for those variables controlled in the randomization process. Differences of varying sizes will exist for the other variables.

35b. *Question:* What if at the end of the study one discovers that an important baseline characteristic was overlooked in the data collection process? Is it reasonable to expect that variable to explain the observed treatment difference?
Answer: No. The expected difference among treatment groups for an unobserved baseline characteristic is the same as that for an observed characteristic, assuming the groups are the product of a properly administered randomization scheme.

Other related questions: 7, 8, 57, 71, and 73.

19.7 QUESTIONS CONCERNING TREATMENT ADMINISTRATION

36. *Question:* What should be done about treatment protocol violations detected during the trial?
Answer: Corrective action should be taken to avoid future violations. The departures noted and actions taken should be reported in publications from the trial.

37. *Question:* Is there a reliable way to measure treatment adherence in drug trials?
Answer: Not really, except in inpatient settings. Various methods have been used to assess drug adherence in studies involving outpatient populations. However, all of them have shortcomings. One method involves use of a tracer substance that is added to the study drugs and that can be assayed in the blood or urine of study patients. One of the shortcomings of this method has to do with formulary problems that arise from the addition of any tracer substance to existing drugs. The choice of substances must be limited to those approved by the Food and Drug Administration and that do not affect the bioavailability or pharmacology of the drugs. Another problem has to do with the mechanics of obtaining blood or urine samples for the adherence test. They are normally collected as part of scheduled follow-up visits. As a result they can provide a biased view of adherence if patients change their medicine-taking behavior in preparation for a forthcoming clinic visit.

Blood or urine tests, designed to detect the presence of the drug itself, can be used when it is not feasible to use a tracer substance. However, results from such tests can be quite variable and may not be specific for the drug. In addition, they suffer from the same problem mentioned above if tests are performed as part of a regular clinic visit.

The advent of miniaturized electronic devices has led to development of electronic pill dispensers that automatically record the times at which medicines are withdrawn from them. Comparison of the observed time record with the one prescribed provides an indirect measure of compliance. Pill counts, based on medications returned to the clinic by the patient, are sometimes used as crude measures of adherence. However, these measures have limited use, especially when patients realize that they are used to check on adherence.

38a. *Question:* Should a patient who either refuses to take his assigned treatment upon entry into the trial or who refuses to continue the treatment after entry be retained in the trial?
Answer: Yes. All patients enrolled in the trial should be retained for follow-up regardless of treatment course.

38b. *Question:* Should patients who are started on their assigned treatment and subsequently found to be ineligible for enrollment be retained for followup?

Answer: Yes, particularly if the assigned treatment is continued. However, even if a treatment change is required the patient should continue to be followed.

39. *Question:* Should patients found to be ineligible for the trial after randomization be continued on treatment?

Answer: The answer depends on the nature of the treatments involved. Obviously, treatment should not be continued if there are contraindications for doing so.

Some study designs require the initiation of treatment before a final assessment of eligibility is made (e.g., a trial involving MI patients who are started on treatment in the emergency room). Treatment may have to be stopped if subsequent tests indicate that the individual did not have the condition under study.

Termination of treatment may not be sensible if the final eligibility assessment occurs some time after the start of treatment and if there is no reason to stop the treatment, as was the case in the UGDP (University Group Diabetes Program Research Group, 1970d).

40. *Question:* Should clinic personnel be provided with a supply of placebo tablets for use in single-masked fashion if it is necessary to stop a patient's assigned treatment temporarily because of a suspected drug reaction in a double-masked trial?

Answer: Single-masked administration of a placebo may be of value when the complaints leading to the termination are vague and there is a desire to determine whether they are due to a real or an imagined cause. The procedure is of less value when the reaction can be documented with laboratory tests or by some other objective means.

The CDP allowed study physicians to use a single-masked placebo on patients who appeared to be having drug reactions (Coronary Drug Project Research Group, 1973a). However, their use created a dilemma for physicians when they were called upon to answer questions from patients concerning their use. Often they were placed in the position of having to tell "white lies" to preserve the mask. The wisdom of this deception is questionable because of the impact it may have on patient-physician relations.

Other related questions: 26, 27, 28, and 29.

19.8 QUESTIONS CONCERNING PATIENT FOLLOW-UP

41. *Question:* Should follow-up of a patient be terminated once he experiences the event of interest?

Answer: No, except when the event itself precludes further follow-up. Added follow-up through the close of the trial for new events can provide additional data for comparison of the treatment groups.

42. *Question:* Is there any way to compensate for losses to follow-up due to dropouts or lack of treatment compliance?

Answer: Yes and no. As noted in Chapter 9, there are ways to increase the sample size to compensate for anticipated losses. However, the increases do not protect against bias if the losses are differential by treatment group.

43. *Question:* Some studies are designed to add a new patient for each one who refuses the assigned treatment, or whenever one drops out. Is this a useful maneuver?

Answer: It can serve the same purpose as the sample size adjustment alluded to in the answer to question 42. However, the practice can lead to a false sense of security if it is perceived as a solution to treatment compliance or dropout problems.

The practice is only useful in preserving the statistical precision of the trial if patient recruitment continues over the entire course of follow-up. It is not a practical means of maintaining the desired type I and II error protection if most of the losses are from patients who drop out after recruitment has been completed.

44. *Question:* Does it pay to try to get patients back under follow-up once they have dropped out?

Answer: Yes, especially in a long-term trial. Periodic contact with patients who have dropped out can be useful in convincing some to resume treatment and to return to active follow-up (see Section 15.3 for further discussion).

45. *Question:* Is it reasonable to assume that patients who remain under active follow-up have the same risk of developing the event of interest as those who do not?

Answer: Often no. Patients who drop out may have different risk factors than those who continue in the study. These differences may place them at a higher (or lower) risk of developing the event of interest.

Other related questions: 5, 38, and 65.

19.9 QUESTIONS CONCERNING THE OUTCOME MEASURE

46. *Question:* Is it all right to use a composite outcome measure as the primary outcome measure for a trial?

Answer: Yes, but it is much better to use a single outcome measure for the primary measure. It is difficult to determine the clinical relevancy of most combinations of outcomes, particularly those due to a mixture of disease processes.

47. *Question:* Should an outcome measure not used in the original sample size calculation, or mentioned in the design documents for the trial, be ignored when results of the trial are analyzed?

Answer: No. All available data should be used in the evaluation of the study treatments. While it is desirable to be as explicit as possible in the design stage regarding the primary outcome measure, failure to designate a variable as an outcome measure does not preclude its use in data analysis. (See Section 20.5 for general precautions.)

48. *Question:* What if the outcome measure is subject to a treatment-related ascertainment bias?

Answer: An effort should be made to assess the nature and magnitude of the bias, and a summary of the problem should be included in the study publication.

Other related questions: 4, 45, 58, 72, 75, and 76.

19.10 QUESTIONS CONCERNING DATA INTEGRITY

49. *Question:* What should be done if someone has tampered with the randomization process?

Answer: The entire set of results from the trial may have to be discarded if the tampering was widespread. The extent of the problem, the way the tampering was done, the way in which it was detected, and the action taken should be reported in the study publication. It should also indicate if the problem led to a data purge and, if so, the amount of data purged. If no purge was made, the paper should indicate why the investigators believe none was required. It is good practice to perform two sets of treatment comparisons when purges involving sizable numbers of patients are made: one set for purged patients and the other set for all remaining patients. The results of the two analyses should be included in a publication from the trial.

50. *Question:* Can exclusion of patients judged to be ineligible after randomization affect the credence placed in the results?

Answer: It can. Elimination of patients who are randomized and subsequently found to be ineligible can bias the results if the judgments on eligibility are made by persons who know the treatment assignments. Exclusions, if allowed at all (see answer to question 39), should be based on data collected before randomization and should be made by individuals masked to treatment assignment.

51. *Question:* What should be done with the data from a clinic in a multicenter trial that withdraws during the course of the trial?

Answer: The answer depends on the reason for the withdrawal. The data should be purged from the database if it was due to questionable data practices. Otherwise they should be retained. Whenever possible, an effort should be made to continue follow-up of patients affected by the withdrawal. Sometimes this can be accomplished by transferring care responsibilities to another clinic, as suggested in the answer to question 2b.

The elimination of data from a clinic will not necessarily have any impact on treatment comparisons, provided the proportionate mix of patients by treatment group in the clinic eliminated is the same as for the remaining clinics.

52. *Question:* Is it possible to change data collection or coding practices during the course of the trial and still have a valid trial?

Answer: Yes, so long as the changes are independent of observed treatment effects. However, it is desirable to minimize these changes for practical as well as scientific reasons.

53. *Question:* What should be done with contrived data?

Answer: The answer depends upon the extent of the problem and on whether the contrivance was treatment related. The results of the entire trial may have to be discarded if the problem is extensive and treatment related, whereas no purge may be required if it is restricted to a few isolated cases.

The Multiple Risk Factor Intervention Trial (MRFIT) elected to retain data from one clinic in which personnel were alleged to have falsified blood pressure data for patients being screened for enrollment (Presberg and Timnick, 1976). On the other hand, the data center in the Eastern Cooperative Oncology Study (ECOG) elected to purge all data contributed by one of its clinics because of the serious nature and extent of the falsification (*Boston Globe*, 1980a, 1980b, 1980c, and 1980d; *Boston Sunday Globe*, 1980).

Manuscripts generated from trials in which data falsification has occurred should indicate the nature of the problem and the action taken, if any, to eliminate the questionable data.

Other related questions: 4.

19.11 QUESTIONS CONCERNING DATA ANALYSIS

54. *Question:* What is the basis for pooling treatment results across clinics in a multicenter trial?

Answer: It stems from the use of common treatment and data collection procedures, and from the ongoing quality assurance procedures designed to detect and minimize procedural differences among study clinics.

55. *Question:* Is randomization required for a valid analysis?

Answer: No. The main purpose of randomization is to provide a method of assignment that is free of selection bias. Randomization theory has been used to form the basis for some tests of significance, but the theory, per se, is not crucial for most of the data analyses carried out in the typical clinical trial.

56. *Question:* Is one obligated to make treatment comparisons in subgroups defined when the trial was designed?

Answer: No. In fact, the first analysis should be without regard to any subgrouping. Secondary analyses may be done within various subgroups, including randomization strata.

57. *Question:* Can differences in the baseline composition of the study groups invalidate treatment comparisons?

Answer: It depends on how large they are and how they occurred. They can if the differences are an expression of a treatment-related bias resulting from a breakdown in the assignment process, but not if they are relatively small and unrelated to treatment.

Much of the discussion concerning the UGDP results published in 1970 (University Group Diabetes Program Research Group, 1970e) centered on the comparability of the treatment groups at the time of randomization. Critics argued that the constellation of baseline entry characteristics present in the tolbutamide-treated patients automatically predisposed them to a higher risk of mortality than was the case for control-treated patients (Feinstein, 1971; Schor, 1971; Seltzer, 1972). Arguments concerning comparability persisted in spite of the fact that the observed differences were within the range of chance, that adjustment for the differences did not materially affect the size of the tolbutamide-placebo difference in mortality, and that analyses by others outside the UGDP reached similar conclusions regarding tolbutamide therapy (Committee for the Assessment of Biometric Aspects of Controlled Trials of Hypoglycemic Agents, 1975; Cornfield, 1971).

58. *Question:* Is it appropriate to consider more than one outcome measure in the analysis of the data?

Answer: Yes. As a matter of fact it is often an essential part of the analysis process. See question 47.

59. *Question:* Are there dangers in analyses that focus simply on patients who received the assigned treatment?

Answer: Yes, they can lead to overestimation of the treatment effect (see Section 18.1).

60. *Question:* Where should data on patients who did not receive the assigned treatment be counted?

Answer: The primary analysis should be based on the original treatment assignment (see Section 18.1). Other analyses, including those based on classification of patients by treatment received, may be carried out.

61. *Question:* How does one take account of changes in a patient's adherence to treatment over the course of the trial?

Answer: The problem with varying levels of adherence is common in drug trials in which patients are expected to remain on their assigned treatment for long periods of time. The primary analysis should be by the initial treatment assignment, without regard to adherence. This analysis can be followed by others that are designed to take account of observed adherence levels (e.g., see University Group Diabetes Program Research Group, 1970e).

62. *Question:* What should be done with data from a patient whose treatment is unmasked for medical reasons?

Answer: They should be analyzed in the treatment group indicated by the randomization. Other analyses may be performed and reported in which data for such patients are excluded to determine if doing so affects the magnitude of the observed treatment effect.

63. *Question:* What if the treatment masking was ineffective? Are the data still worth analyzing?

Answer: Masking is never 100% effective. Treatment-related side effects may reveal the treatment assignment to both patients and physicians. The validity of treatment comparisons will depend on whether or not the deficiencies in masking allowed introduction of treatment-related biases.

64. *Question:* What should be done with data for patients whose treatment assignment was needlessly unmasked?

Answer: The analysis approach should be similar to that outlined for question 61. However, the frequency of frivolous unmaskings should be noted in the published report. A large number may be indicative of a lack of regard for the study protocol by investigators in the trial and may raise general questions regarding the validity of the study.

65. *Question:* How does one deal with missing data caused by losses to follow-up?

Answer: While there is no substitute for complete follow-up, the usual approach is to carry out a series of analyses, each requiring a different set of assumptions regarding the rate of outcome events after patients are lost to follow-up. One of the analyses should be done assuming a zero event rate over the periods patients are lost to follow-up. Other analyses may be done in which all patients lost to follow-up are assumed to have had the event after loss to follow-up, or alternatively, in which they are assumed to have experienced the event at the same rate as a defined portion of the study population (e.g., the control-treatment group of patients who remained under active follow-up). Losses are not a serious source of concern if the various analyses all support the same basic conclusion and if they are not differential by treatment group.

66. *Question:* How should aberrant laboratory results be handled?

Answer: Outlier values, whether they are a legitimate indicator of some underlying biological problem or are due to a laboratory or recording error, may have to be trimmed or eliminated in analyses involving means or variances. The rules for trimming or elimination should be constructed and administered without regard to treatment assignment or effect and should be specified in published reports from the trial.

67. *Question:* What if there is a secular trend in the laboratory data generated in a trial? Will this affect comparisons between treatment groups?

Answer: It should not, assuming that patients in all treatment groups were enrolled over the same time frame and that the time sequence in which laboratory determinations were performed was independent of treatment assignment.

68. *Question:* How should data obtained from interim unscheduled examinations be handled?

Answer: The first analysis should be done ignoring the results. A second one may be done with the results included. A differential rate of interim unscheduled examinations by treatment group can influence the rate at which nonfatal events are diagnosed and reported. CDP investigators were sufficiently concerned about this possibility as to virtually ignore results from unscheduled examinations when analyzing the dextrothyroxine results (Coronary Drug Project Research Group, 1972; 1981).

69. *Question:* Is it permissible to perform analyses during the course of the trial to detect treatment effects?

Answer: Yes. They are not only permissible but required in any trial in which the treatments are hazardous, or in which early detection of a treatment effect may prove beneficial to patients already in the trial or to those yet to be enrolled (see Chapter 20).

70. *Question:* What if there is a major time lag in the flow of data from the clinic to the data center? Can this have an impact on the detection of treatment differences during the trial?

Answer: Yes, especially if the time lag is differential by treatment group. Procedures should be established to ensure data flows that are timely and uniform with regard to treatment assignment (see Chlebowski and co-workers, 1981).

71a. *Question:* Is it reasonable to argue that imbalance in the distribution of an important but unobserved baseline risk factor could account for an observed treatment difference or lack of one in a randomized trial?

Answer: Not really. As noted in the answers to questions 35a and 35b, the expected distribution of an unobserved characteristic is the same as for an observed characteristic.

71b. *Question:* Is a trial invalid if there are differences among the treatment groups with regard to key baseline variables?

Answer: Generally no, unless the differences are due to selection biases arising from a breakdown in the way treatment allocations were made.

72. *Question:* Is it appropriate to use a subset of deaths as the prime outcome measure?

Answer: The trial may be designed for detection of a specified difference for a subset of deaths, as was the case in MRFIT (Multiple Risk Factor Intervention Trial Research Group, 1982). However, the initial analysis should be for mortality from all causes (see question 75b).

Other related questions: 4, 8, 34, 35, 45, 47, 48, 51, 53, 75, and 76.

19.12 QUESTIONS CONCERNING CONCLUSIONS

73. *Question:* Is it really possible to draw any conclusions from a clinical trial because of the select nature of the study population involved?

Answer: Yes. Comparisons between treatment groups are valid so long as all groups have been exposed to the same selection factors.

74. *Question:* Is it possible to generalize findings beyond the population studied and the treatments used?

Answer: Any generalization that goes beyond the study population must be made with caution and is judgmental rather than statistical in nature. Treatment effects observed in a specified population with a particular dosage of a drug may not be generalizable to a broader population. Similarly, an effect produced with one formulation of a compound may not be produced by a sister product. For example, it is tempting to generalize the UGDP findings on tolbutamide to other sulfonylurea compounds. However, the study included only one member of the family (University Group Diabetes Program Research Group, 1970d). The question of scientific validity versus generalizability is touched upon by the National Diet-Heart Study Research Group (1968).

75a. *Question:* Is it appropriate to base conclusions from a trial on a nonfatal event if there is differential mortality by treatment group?

Answer: No. Conclusions based on differences in a nonfatal outcome are only valid if there is no difference among the study groups with respect to mortality. A differential mortality by treatment group may influence the rate of occurrence of nonfatal events. The treatment group with the highest mortality rate may have the lowest nonfatal event rate if death occurs before patients have a chance to develop the nonfatal event of interest.

75b. *Question:* Is it appropriate to base conclusions on deaths due to a specific cause (e.g., cardiovascular deaths)?

Answer: Only if the conclusion is consistent with the one reached when all deaths are considered.

76. *Question:* Is it appropriate to base conclusions on an outcome measure that was not

expected to yield a difference when the trial was designed?

Answer: Yes, especially when the measure has more clinical relevance than the one used in the design of the trial. The focus on mortality in assessment of the tolbutamide and phenformin results in the UGDP, even though the study was designed to look for differences in nonfatal outcomes, is a case in point (University Group Diabetes Program Research Group, 1970d, 1970e, 1975).

Other related questions: 10, 26, 52, 54, 55, 57, 71, and 72.