1
Randomization and the Clinical Trial

1.1 INTRODUCTION

The goal of any scientific activity is the acquisition of new knowledge. In empirical scientific research, new knowledge or scientific results are generated by an investigation or study. The validity of any scientific results depends on the manner in which the data or observations are collected, that is, on the design and conduct of the study, as well as the manner in which the data are analyzed. Such considerations are often the areas of expertise of the statistician. Statistical analysis alone is not sufficient to provide scientific validity, because the quality of any information derived from a data analysis is principally determined by the quality of the data itself. Therefore, in the effort to acquire scientifically valid information, one must consider all aspects of a study: design, execution, and analysis.

This book is devoted to a time-tested design for the acquisition of scientifically valid information – the randomization of study units to receive one of the study treatments. One can trace the roots of the randomization principle to Sir R. A. Fisher (e.g., 1935), the founder of modern statistics, in the context of assigning “treatments” to blocks or plots of land in agricultural experiments. The principle of randomization is now a fundamental feature of the scientific method and is employed in many fields of empirical research. Much of the theoretical research into the principles and properties of randomization has been conducted in the domain of its application to clinical trials. A clinical trial is basically an experiment designed to evaluate the beneficial and adverse effects of a new medical treatment or intervention. In a clinical trial, often subjects sequentially enter a study and are randomized to one of two or more study treatments. Clinical trials in medicine differ in many respects from randomized experiments in other disciplines, and clinical trials in humans involve complex ethical issues, which are not encountered in other scientific experiments. The use of
randomization in clinical trials has not been without controversy, as we shall see, and statistical issues for randomized clinical trials can be very different from those in other types of studies. Thus this book shall address randomization in the context of clinical trials.

Randomization is an issue in each of the three components of a clinical trial: design, conduct, and analysis. This book will deal with all three elements; however, we will focus principally on the statistical aspects of randomization in the clinical trial, which are applied in the design and analysis phases. Other, more general, books are available on the proper conduct of clinical trials (see, e.g., Tygstrup, Lachin, and Juhl, 1982; Buyse, Staquet, and Sylvester, 1984; Pocock, 1984; Piantadosi, 2013; Friedman, Furberg, and DeMets, 2010; Chow and Liu, 2013; Matthews, 2006). These references also give a less detailed development of randomization.

1.2 CAUSATION AND ASSOCIATION

Empirical science consists of a body of three broad classes of knowledge: descriptions of phenomena in terms of observable characteristics of elements or events; descriptions of associations among phenomena; and, at the highest level, descriptions of causal relationships between phenomena. The various sciences can be distinguished by the degree to which each contains knowledge of the three classes. For example, physics and chemistry contain large bodies of knowledge on causal relationships. Epidemiology, the study of disease incidence, its risk factors, and its prevention, contains large bodies of knowledge on phenomenologic and associative relationships.

Although a major goal of epidemiologists is to determine causative relationships, for example, causal relationships between risk factors and disease that can potentially lead to disease prevention, the leap from association to causation is a difficult one. Jerome Cornfield’s (1959) treatise on “Principles of Research” gives a beautifully written account of the history of biomedical studies and the emergence of principles underlying epidemiological research.

Cornfield points to a mass inoculation against tuberculosis in Lübeck, Germany, in 1926. A ghastly episode occurred where 249 babies were accidentally inoculated with large numbers of virulent bacilli. In a follow-up of those babies, 76 had died, but 173 were still free of tuberculosis when observed 12 years later. If the tuberculosis bacilli cause tuberculosis, why didn’t all the children develop the disease? The answer, of course, is the dramatic variability in human response to even large doses of a deadly agent. Thus, as we all know, tuberculosis bacilli cause tuberculosis, but causation in such cases, it does not mean that all those exposed to a pathogen will experience the ill effects.

Similarly, one can ask the famous question, why doesn’t everyone who smokes develop lung cancer? One possible answer that would please the tobacco industry is that there is a hormonal imbalance that both causes lung cancer and causes an insatiable craving for cigarettes. An alternative answer is that there are competing risks: something else kills them first. The most probable answer is that not all those who smoke will develop cancer, due to biological or genetic variation.
Humans have such a complex and varied physiology; they are exposed to so many different environmental conditions; their health is also deeply tied to complex mental states. How can a scientist possibly sift through all the associations one can find between health and these other factors to find causes or cures for disease? One of the oldest principles of scientific investigation is that new information is obtained from a comparison of alternate states. Thus, a controlled clinical trial is an experiment designed to determine if a medical innovation (e.g., therapy, procedure, or intervention) alters the course of a disease by comparing the results of those undertaking the innovation with those of a group of subjects not undertaking the innovation.

Perhaps the first comparative study of record is the biblical account of Daniel (Chapter 1) in approximately 605 BCE, on the effects of a vegetarian diet on the health of Israelites. Rather than be placed on the royal diet of food and wine of the Babylonian court, Daniel requested that his people be placed on a diet of vegetables.

“Test us for ten days,” he said, “... then compare us with the young men who are eating the food of the royal court, and base your decision on how we look....”

When the time was up, they looked healthier and stronger than all those who had been eating the royal food.

Another famous example of a controlled intervention study is Lind’s (1753) account of the effects of different elixirs on scurvy among British seamen. His study showed the beneficial effects of citrus and led (50 years after the study) to the Royal Navy’s decision to store citrus on long voyages.

While the idea of comparing those on the innovative treatment with a control group sounds obvious to us today, historically it was not always entirely clear whom to include in the innovation and control groups. At the turn of the twentieth century, an antityphoid inoculation movement created controversy between Sir Almroth Wright, a famous immunologist, and Karl Pearson, who, along with Fisher, was a founder of modern statistics. Sir Wright gave the inoculation to anyone who wanted it and compared the subsequent incidence of typhoid with a group of men who refused the inoculation. Here is Pearson’s first writing on the subject (Cornfield, 1959, pp. 244–245):

Assuming that the inoculation is not more than a temporary inconvenience, it would seem possible to call for volunteers, but while keeping a register of all men who volunteered only to inoculate every second volunteer. In this way any spurious effect really resulting from a correlation between immunity and caution would be got rid of.

Four years later, Pearson’s opinion was even stronger:

Further the so-called controls cannot be considered true controls, until it is demonstrated that the men who are most anxious and particular about their own health, the men who are most likely to be cautious and run no risk, are not the very men who will volunteer to be inoculated. ... Clearly what is needed is the inoculation of one half only of the volunteers, equal age incidence being maintained if we are to have a real control.
Pearson recognized what the immunologist did not: that human response to infectious, preventive, or therapeutic agents is variable and is positively related to patient characteristics, such as a willingness to volunteer to receive a new treatment. Thus, positive steps must be taken in the design and conduct of a study to eliminate sources of incomparability between those treated and the controls. The inoculated group cannot be compared to any arbitrary control group. The control group must be comparable to the treated group with respect to immune background, hygiene, age, and so on. Such factors are called confounding variables, because incomparability of the groups with respect to any such factors may confound the results and influence the answer to the research hypothesis.

These considerations play a major role in the design, conduct, and analysis of epidemiologic studies today. In an observational epidemiologic study, naturally occurring populations are studied to identify factors associated with some outcome. Since such studies do not employ a randomized design, the results are subject to various types of bias (cf. Breslow and Day, 1980, 1987; Rosenbaum, 2002; Selvin, 2004; Kelsey, et al., 1996; among others). In a retrospective study, these populations consist of cases that develop the disease and controls that do not, so that a direct comparison can be made. Just as Pearson noted that there should be equal age incidence in both the inoculated and control groups, epidemiologists may also use matching on important variables (covariates or prognostic factors) that may confound the outcome. Matching is usually done, for instance, on important demographic factors, such as gender, age, and race. Each “case subject” will have a “control subject” with similar characteristics on matched covariates. This allows for greater comparability between the comparison groups. However, it is impossible to match on all known covariates that may influence outcome. Therefore, the leap from association to causation is again tenuous.

The most famous epidemiologic studies were those that demonstrated that smoking causes lung cancer. In 1964, the Report of the Advisory Committee to the Surgeon General was issued that led to warning labels on cigarette packages and restrictions on advertising. The report summarized the evidence from numerous studies that had shown an association between smoking and increased risk of lung cancer and other diseases. Despite any randomized controlled experiments, and based only on observational studies, the Committee concluded that the epidemiologic evidence showed that smoking was indeed a cause of lung cancer. The establishment of a causal relationship between tobacco smoking and cancer created much controversy (and does to this day in some circles). The Surgeon General’s report on “The Health Consequences of Smoking” clarified the issue with a definitive statement on what types of evidence from observational studies can lead to a determination of a causal relationship. The Committee (1982, p. 17) stated:

The causal significance of an association is a matter of judgment which goes beyond any statement of statistical probability (sic).... An entire body of data must exist to satisfy specific criteria;... when a scientific judgment is made that all plausible confounding variables have been considered, an association may be considered to be direct (causal)....
The Committee stated that the following five criteria must be satisfied:

1. **Consistency of the association.** Diverse methods of approach should provide similar conclusions. The association should be found in replicated experiments performed by different investigators, in different locations and situations, at different times, and using different study methods.

2. **Strength of the association.** Measures of association (e.g., relative risk, mortality ratio) should be large, indicating a strong relationship between the etiologic agent and the disease.

3. **Specificity of the association.** Specificity refers to the precision with which one component of an associated pair predicts the occurrence of the other component in the same individual. For instance, how precisely will smoking predict the occurrence of cancer in an individual? The researcher must consider that agents may be associated with multiple diseases and that diseases may have multiple causes. A single naturally occurring substance in the environment may cause the disease. A single factor can also be a vehicle for several different substances (e.g., tar and nicotine in tobacco), and these may have synergistic or antagonistic effects. There is also no reason to believe that one factor has the same relationship with a different disease with which it is associated. For example, smoking is also associated with heart disease, but perhaps in conjunction with dietary factors that are not important in lung cancer.

4. **Temporal relationship of the association.** Exposure to the etiologic agent must always precede the disease.

5. **Coherence of the association.** The association must make sense in light of our knowledge of the biology and natural history of the disease.

The nine largest studies cited in the Surgeon General’s report comprised almost 2 million patients with 17.5 million patient-years of exposure. Based on these data and the convergence of evidence from other sources, one can be confident that smoking “causes” lung cancer, even though the precise causal agent has not been identified (i.e., tar, nicotine, or other agents) and even though no randomized experiment of the effects of smoking and lung cancer has ever been performed.

The overriding question in determining causality in such instances is whether the design or analysis has controlled or “adjusted” for all possible extraneous variables that might account for higher incidence of the disease. Some would say that only a randomized study can ensure adequate control for such factors. However, randomized studies of risk factors such as smoking are impossible, in that it is unethical to randomly assign risk factors to patients. It is instructive to note that Fisher, the father of randomization, was never convinced of the link between smoking and lung cancer, and perhaps equally instructive to note that he was a dedicated smoker. Today, almost all epidemiologists and biostatisticians will accept consistent, replicated, careful observational evidence, and few would argue the potency of the evidence against tobacco.

However, it is rare that an adequate body of evidence is amassed from epidemiologic studies alone to assert the aforementioned conditions. The number of studies, patients, and extent of exposure required to establish a definite cause by
epidemiologic investigation are far greater, and the results ultimately less compelling, than those obtained from a randomized clinical trial, when such trials are possible.

1.3 RANDOMIZED CLINICAL TRIALS

In this book, we will refer to clinical trials that are prospective comparisons of two or more treatments, one or more of which is a new innovation under test, and one or more of which is a control. The most common is a therapeutic trial, in which a new therapy, such as a pharmaceutical agent (drug), is compared to a conventional therapy. In a placebo-controlled clinical trial of a new pharmaceutical agent, a group of drug-treated subjects may be compared to a group who receive a placebo control [a placebo being a drug preparation (e.g., pill) that is identical to the active therapy, but with inert (inactive) ingredients]. When an established therapy already exists, the new drug may be compared to an active control, where the control group receives the established therapy. Therapeutic pharmaceutical clinical trials are often called phase III clinical trials, because they represent the third phase of a four-phase process in investigating a promising new therapy. From development of a new pharmaceutical agent to its approval, there is often a phase I clinical trial, a small trial to determine the potential toxicity of different dose levels of a drug, and a phase II clinical trial, a preliminary study of toxicity and efficacy. A phase IV clinical trial involves post-approval follow-up of patient status. These phases are particularly seen in the study of cancer chemotherapeutic agents (see Buyse, Staquet, and Sylvester, 1984), and the four-phase process is often streamlined in other specializations of medicine.

The innovation, however, need not be a simple drug. In some cases, a new procedure is evaluated. An example is the Lupus Nephritis Collaborative Study that desired to assess the effects of plasmapheresis on the progression of lupus nephritis or kidney disease associated with lupus erythematosis (Lewis, et al., 1992). Patients in the plasmapheresis group were hospitalized for a month to undergo daily plasma filtration and exchange, followed by the initiation of standard immunosuppressive therapy consisting of cytoxan and prednisone, the dose of the latter tapered when the patient responded favorably. The patients in the control group received comparable immunosuppressive therapy without initial plasmapheresis.

In other cases, a new intervention or an entire treatment regimen of multiple therapies is compared to a control regimen. An example is the Diabetes Control and Complications Trial, which was designed to assess whether a program of intensive therapy aimed at maintaining near-normal levels of blood glucose in subjects with type I diabetes mellitus would prevent or retard the progression of microvascular complications associated with diabetes. Patients in the intensive treatment group received aggressive insulin therapy with frequent monitoring of glucose levels, in conjunction with dietary counseling and exercise. Patients in the conventional treatment group received conventional therapy aimed at maintaining general well-being. While intensive therapy greatly reduced the risks of complications compared to conventional therapy, such an overall comparison alone cannot identify the mechanism by which the treatment had
its effects (Diabetes Control and Complications Trial Research Group, 1993). Subsequent analyses, however, indicated that the effects of intensive treatment were indeed wholly accounted for by the reductions in blood glucose levels.

Some call such trials pragmatic trials because the innovation consists of two or more possible agents or procedures used in combination, such that the overall group comparisons alone cannot identify the mechanism by which the innovation produces its effects. However, the pragmatist would argue that conclusive evidence that the innovation is indeed beneficial in practice is adequate for its adoption even when the mechanism of the effect is unknown.

The pivotal component of phase III clinical trials is randomization or random assignment of patients to receive either the experimental treatment(s) or control. Cornfield (1959, p. 245) summarized the importance of randomization:

1. It controls the probability that the treated and control groups differ more than a calculable amount in their exposure to disease, in immune history, or with respect to any other variable, known or unknown to the experimenter, that may have a bearing on the outcome of the trial. This calculable difference tends to zero as the size of the two groups increase.
2. It makes possible, at the end of the trial, the answer to the question “In how many experiments could a difference of this magnitude have arisen by chance alone if the treatment truly has no effect?” It may seem mysterious that a mathematician could actually predict the course of future experiments. All you have to do is compute what would happen if a given set of numbers were randomly allocated in all possible ways between the two groups. Randomization allows this.

The first property of randomization is that it promotes comparability among the study groups. Such comparability can only be attempted in observational studies by adjusting for or matching on known covariates, with no guarantee or assurance, even asymptotically, of control for other covariates. Randomization, however, extends a high probability of comparability with respect to unknown important covariates as well. The second property is that the act of randomization provides a probabilistic basis for an inference from the observed results when considered in reference to all possible results. This randomization approach to inference is very different from the usual testing of unknown parameters arising from an independent and identically distributed sample from a known distribution. Later we will deal in detail with these and other precise statistical properties of randomization.

In Cornfield’s first point, we come to the root importance of the randomized clinical trial. As scientists are interested in descriptions of phenomena, association among phenomena, and then mechanisms of causation, then the biomedical studies for each require increasing standards of evidence. Basic science research often involves the description of phenomena, observational studies lead to the determination of associations among phenomena, and randomized clinical trials lead to definitive statements on causative effects of agents or regimens on disease processes. As we have seen, despite the fact that consistent, replicated observational studies can also lead us to determine causality, there may always be questions as to whether we have controlled
for all factors relating to incidence and prognosis of a disease. The randomized clinical trial allows this control and, hence, represents the highest standard of evidence among biomedical studies.

Among the first clinical trials, as we know them today, were the trials performed under the direction of Sir Bradford Hill in the 1940s by the Medical Research Council. These were the first medical trials to employ randomization to individual patients and constituted a major advance. They led to important findings in many of the persistent diseases of the day, such as whooping cough and tuberculosis. In every respect, they were similar to the most rigorous trials conducted today.

The polio vaccine trial of 1954 changed the face of public health worldwide (see Francis, 1955). Approximately 400,000 children were randomized to receive either the vaccine or a saltwater injection. The results showed a relative risk of 2.5, in favor of the vaccine group. The success of this study belies the controversy among study participants about the need for a controlled, randomized study. In fact, in a quotation attributed to Jonas Salk, it appears that Salk was not convinced of the need for a placebo control in the polio trial (source unknown):

> In talks with many people in our own group … and others as well, I found but one person who rigidly adhered to the idea of a placebo control and he is a bio-statistician who, if he did not adhere to this view, would have had to admit his own purposelessness in life.

In the end, randomized controls were felt necessary because of the variability of incidence of polio from year to year. It was largely due to trials like the polio vaccine trial that convinced the medical community of the value of the randomized clinical trial. Today, it is often considered the “gold standard” among techniques of ascertaining medical evidence.

A good example of the benefits of randomization can be seen in the National Cancer Institute’s clinical trial of 62,000 women covered by the health insurance plan (HIP) of Greater New York, commonly known as the HIP trial (see Cairns, 1985). The women were randomized into a “test” group, who were offered a free annual physical examination and mammography for early detection of breast cancer and a “control” group who were given no special encouragement to be examined. The trial was designed to determine if the act of offering free mammography examinations reduces deaths from breast cancer. The results were encouraging. Among the test group, there were 2.9 deaths per 1,000 women in the first 9 years, and among the control group, there were 4.1 deaths per 1,000 women. The two groups were comparable in their incidence of breast cancer and in terms of general mortality from causes other than cancer, as should be the case because the experiment was randomized. But the results of the trial were also interesting because, among the test group, those who refused examination had a lower death rate due to breast cancer (2.8 per 1,000) than those who accepted the mammography (3.0 per 1,000). This demonstrates the danger of accepting observational data at face value, as one might have concluded that mammography was not effective. The acceptance and rejection groups within the
test group were self-selected and, hence, subject to confounding due to incompara-
bility with respect to important covariates. In this case, Cairns (1985) believes the
confounding variable to be education level. Since better-educated women are known
to be more likely to have breast cancer, and less well-educated women are more likely
to have less interest in their health, and consequently are more likely to reject exami-
nation, the observational component of this study was biased in favor of the rejection
group (for an instructive set of homework problems on the HIP data, see Freedman,
Pisani, and Purves (1998), Problems 9 and 10, pp. 22–23.)

1.4 ETHICS OF RANDOMIZATION

Randomized clinical trials use probability as a method of assigning treatments to
patients. Many have argued that probability has no role in medicine and that only a
physician can decide which treatment a patient should receive, using his or her best
judgment. However, clinical trials present a unique situation in which new innova-
tions, such as investigational drugs, are being tested for efficacy and safety. Until a
drug is proven to be effective and adequately safe, or ineffective or harmful, or just
ineffective, the physician is in a state of equipoise: a state of genuine uncertainty
about which experimental therapy is more effective. Most ethicists would agree, in
principle, with the concept that it is ethical to employ randomization in a state of true
equipoise, provided the patient consents to be a study participant and is fully informed
about the potential benefits and risks of the treatments to be compared in the study.

However, ethics involving human experimentation are seldom so simplistic. On the
one hand, a clinical trial gives the patient a chance of being assigned to a potentially
beneficial therapy that would not be obtainable elsewhere. But that therapy may also
be highly toxic. There is also a chance that a patient will be assigned to a placebo,
in effect being denied a therapy that may later prove to be very beneficial (or, on the
other hand, harmful). Decisions to enroll in a clinical trial are difficult ones, for this
reason, and the patient must often be willing to make a sacrifice for the benefit of our
public health.

These considerations exemplify the delicate balance between individual ethics and
collective ethics (see Palmer and Rosenberger, 1999). Individual ethics dictate what
is best for the individual patient, while in collective ethics, we consider the advance-
ment of public health through careful scientific experimentation. In a broad sense,
collective ethics leads to individual ethics, as it is only when careful scientific experimen-
tation has yielded a universal standard of care for a given disorder that physicians
will be fully informed and will have a scientific basis for the assignment of the best
therapy to an individual patient. Although experimentation may lead to many patients
being assigned an inferior therapy prior to the determination of the standard of care,
this is the price an informed society must pay to obtain the evidence necessary to
support informed therapeutic decisions. Such ethical dilemmas are naturally contro-
versial and are the subject of many treatises and texts (e.g., Engelhardt, 1996).

Some would argue that equipoise is rarely present at the beginning of a phase III
clinical trial. Animal studies and phases I and II clinical trials data, plus information
on the biological action of the innovation (e.g., drug), combine to create in the mind of many physicians a belief in the effectiveness of one therapy over another. But such confidence may often be premature. The literature is replete with results of negative studies, where promising therapies were shown to be ineffective or even terribly harmful. If equipoise is defined in the confines of a single physician’s “hunches” or intuition about a therapy rather than in a global standard of evidence based on randomized controlled studies, there will be no advancement of medical science. This is not to say that careful, replicated, consistent observational studies, as defined in Section 1.2, are not useful and cannot be convincing. But randomization adds an additional component that mitigates contention, and the National Institutes of Health and U.S. Food and Drug Administration now consider a well-conducted, randomized clinical trial to be of vital importance in demonstrating the efficacy and safety of a new therapy. As Cassel (2009) points out in the *Proceedings of the National Academy of Sciences*,

While the ideal of ethics posits equipoise, the reality is a struggle in the minds of both the investigators and the patients themselves. The investigators are told not to posit the “new” treatment as “better,” and yet the patients coming into the trial often agree to the trial precisely because they hope to find a better treatment for their condition. . . . The policy prescription for this problem was the requirement for informed consent, with clear guidelines about what kinds and levels of information were to be provided and what attributes of setting and patient condition constituted adequate freedom from undue influence for valid consent. Federal regulations established policy for federally funded research, and a substantial and costly infrastructure of internal review boards ([IRBs]) was established, both academic and free-standing. The scope of this ethics enterprise grew as the scope of the clinical trials industry grew, to the point where there are 4000 IRBs registered with the federal government . . . and > 26,000 clinical trials were published in 2008 alone . . .

Some have also argued that randomized controls are unnecessary and unethical in studies where there are some data already available on the natural history and progression of the disease studied. Rather, they propose that a current cohort of experimentally treated patients might just as well be compared with a past cohort of patients receiving an earlier or no treatment, that is, a cohort of historical controls. In cases where one observes a complete dramatic reversal of the course of a disease, such as the effects of penicillin on a bacterial infection, such evidence may be convincing. However, most therapies yield modest effects and historical controls are subject to various biases that may easily skew the study results. The basic problem is that the historical control group might have very different characteristics from experimental cohort that may bias the study. Such factors might include patient selection criteria, diagnostic methods, the nature of follow-up observations, the criteria for response, and the extent of administration of concomitant medications. A difference between groups in any one of these factors or other factors could result in differences between groups with respect to study outcomes.
While most of today’s scientists have embraced the randomized clinical trial, occasionally particular clinical trials arise that elicit passionate opposition on ethical grounds. A prime example is the recent clinical trials program in third-world countries on the benefits of short-term zidovudine (AZT) therapy in reducing maternal–infant HIV transmission. In a landmark clinical trial, Connor et al. (1994) show that 6 weeks of AZT therapy in pregnant women with HIV reduced the transmission to the infant by two-thirds. The results of this trial were hailed in the medical community, and 6 weeks of antiretroviral therapy quickly became the standard of care for HIV-positive pregnant women in the United States. Unfortunately, the prohibitive cost of zidovudine has prevented developing countries from implementing what is now the standard regimen in the United States. Consequently, a large group of scientists determined that clinical trials should be conducted in these countries using a shorter, less costly regimen of antiretroviral therapy, and such trials were begun with funding from the U.S. government. In an editorial, Lurie and Wolfe (1997) argue that placebo-controlled trials in developing countries are unacceptable, since an effective therapy had already been found in the United States:

… On the basis of the [Connor et al. data], knowledge about the timing of perinatal transmission, and pharmacokinetic data, the researchers should have had every reason to believe that well-designed shorter regimens would be more effective than placebo. These findings seriously disturb the equipoise … necessary to justify a placebo-controlled trial on ethical grounds.

In addition, they argue that, since the standard of care in developing countries (i.e., not providing therapy) is not based on consideration of alternate treatments or clinical data, and rather is based on economic considerations, researchers have an ethical responsibility to provide treatment that conforms with the standard of care in the sponsoring country (i.e., the United States).

This editorial led to much debate in the medical literature. Several of the researchers on these clinical trials in developing countries responded with their own editorial (Halsey et al., 1997). They argue that a placebo control arm is necessary in order to determine if the short course of zidovudine is effective in these countries. Furthermore, they state that providing the same level of care routinely provided to mothers and their infants in the United States would violate the guideline to avoid undue inducements for participation in research and would make the research totally impractical.

If these unsustainable services were provided on a temporary basis what would happen when the research project ended and local practitioners could no longer provide diagnostic tests, infant monitoring, and intensive care units necessary to support the regimen?

They close by noting that many dramatic interventions in developing countries could have been prevented had such “medical and ethical imperialism” been imposed on participants in international studies.
The Declaration of Helsinki was revised in October 2000, adding the following statement:

The benefits, risks, burdens and effectiveness of a new method should be tested against those of the best current prophylactic, diagnostic, and therapeutic methods.

Although this does not exclude the use of placebo in studies where no proven prophylactic, diagnostic, or therapeutic method exists, this new directive is very controversial since some interpret it to mean that a placebo should never be used whenever effective therapy is available.

At issue, however, is not the act of randomization but rather the choice of the control treatment, either placebo or an active control (when the latter exists). Randomization of treatments to patients is now considered the seminal element of a clinical trial for the evaluation of a new innovation in medical care. The purpose of this book is to describe the theoretical basis for the various types or approaches to randomization commonly employed, to describe their statistical properties, and to describe considerations in their practical implementation.

1.5 PROBLEMS

1.1 From a recent issue of any major medical journal (e.g., New England Journal of Medicine, Journal of the American Medical Association), select an article that presents results of a controlled clinical trial involving at least 50 patients. The study should focus on a clinical result (i.e., effectiveness or safety of a treatment) rather than physiologic results (e.g., laboratory or physical measurements).

(i) Give a detailed description of the study design.
(ii) Provide a critique of the study design in regard to the potential for bias in the study results or conclusions. Did the authors describe the choice of study design well and describe possible pitfalls of the design?
(iii) Based on this study, if you were the statistician for a new study (either for a new treatment for the same disease or a study confirming results of the study), describe how you would design a study using randomized controls.
(iv) Alternatively, describe how you would design a study using nonrandomized controls.
(v) Discuss the implications for a randomized versus a nonrandomized study on the interpretation of the results. Which would be preferable?

1.2 From a recent issue of a medical or epidemiologic journal, select an article that presents the results of a nonrandomized observational study of a risk factor associated with an increase or decrease in the risk of a disease or adverse disease outcome.

(i) Give a detailed description of the study design.
(ii) Provide a critique of the study design in regard to the potential for bias in the study results or conclusions. Did the authors describe the choice of study design well and describe possible pitfalls of the design? Which possible biases are cited by the authors and what steps were taken, if any, to address them? Can you identify other possible sources of bias?

(iii) Based on this study, if you were the statistician for a new study (for either a new treatment for the same disease or a study confirming results of the study), describe how you would design a study using randomized controls, if possible.

(iv) Alternatively, describe how you would design a study using nonrandomized controls.

1.3 If you were the statistician on a steering committee, which is deciding whether to participate in a placebo-controlled clinical trial of maternal–infant HIV transmission and short-term AZT in a developing country, where the country has no access to the standard-of-care therapy in the United States (i.e., long-term AZT therapy), what would your stance be? Prepare a 5 minute position paper for a classroom debate. You are asked to respond to the following questions:

(i) Are such trials necessary and ethical?
(ii) Should any placebo-controlled study be adopted?
(iii) Are studies with historical controls reasonable?
(iv) What are the alternatives?

1.4 Are the considerations of individual and collective ethics the same in all clinical trials? Suppose you cross-classified a disease with respect to severity and incidence. For instance, you could have a 4-by-4 table with ordinal categories ranging from 1 to 4. For severity, the categories could range from 1 = mild to 4 = life-threatening. Similarly, incidence could range from 1 = very rare to 4 = very common. Within each cell of the cross-classification, determine the relative importance of individual versus collective ethics. (Palmer and Rosenberger, 1999)

1.6 REFERENCES


