Large and Simple Randomized Trials

Background Paper 3

SUMMARY

For addressing many important research questions, randomized trials are neither necessary nor desirable. However, if the effects of a hypothesized intervention are likely to be only small to moderate in size, a randomized trial with a large sample size will be necessary to provide a definitive test of such research questions. Large trials, if properly designed, can be conducted using relatively simple protocols in which minimal screening or data collection is required.

Large and simple trials are characterized by their emphasis on enrolling large numbers of participants; testing an intervention's effect on a readily ascertained, clinically important outcome; and collecting a relatively limited amount of baseline and followup data. Such trials are particulary appropriate for addressing questions about the relative effectiveness of treatments with wide potential applicability. Because they enroll such a broad range of participants, their results are directly relevant to the wide range of patients seen in clinical practice. Because such trials often involve nonacademic as well as research-oriented clinicians and health care institutions, their results also may be more rapidly incorporated into standard care of patients.

Not all areas of medical research are suitable for large, simple trials. Nevertheless, many questions could be tested using far simpler protocols than those that have been used in most randomized clinical trials. Where appropriate, large and simple trials can provide more reliable tests of an intervention than can other feasible research approaches, and do so at very low cost per patient randomized.

by

Julie E. Buring,
Michael A. Jonas, and
Charles H. Hennekens

Division of Preventive Medicine, Department of Medicine Brigham and Women's Hospital and Department of Ambulatory Care and Prevention
Harvard Medical School
Boston, MA

Two types of epidemiologic studies can be used to test hypotheses: observational studies (case-control or cohort investigations) and randomized trials. Because of the methodologic limitations inherent in observational studies, randomized controlled trials (in which the investigators allocate the treatment to participants at random) represent the type of analytic study in humans that most closely resembles the highly controlled experiments possible in the laboratory (29).

Randomized controlled trials are particularly useful for detecting small to moderate effects of treatments or interventions-effects that are likely to change outcomes by 10 to 50 percent. In such circumstances, observational studies, which evaluate self-selected exposures and the subsequent occurrence of disease, are particularly vulnerable to the effects of unmeasured or unmeasurable confounding factors that may account for all or part of any observed association. For example, in an observational study that reports a 30-percent lower risk of cancer among individuals with high dietary intake of the antioxidant vitamin beta-carotene, the participants with greater intake of this micronutrient might have other dietary or lifestyle practices not fully accounted for in the study analysis that might be partially or entirely responsible for the observed benefit.

Even when an observational study reports a large effect, the amount of uncontrolled confounding may affect the magnitude of the estimated relative risk. Confounding factors, for example, could mean that the reported 15- to 20-fold higher risk of lung cancer among lifelong smokers than among nonsmokers could actually be as high as 25 or as low as 10. It is unlikely, however, that the confounding factors would change the conclusion that a strong relationship exists between smoking and lung cancer. In the case of current smoking and coronary heart disease, if the true effect of smoking is about an 80-percent increased risk of heart disease, uncontrolled confounding may mean that the observed effect is as small as 60 percent or as large as 100 percent. Again, however, this uncertainty does not materially affect the conclusion that current cigarette smoking increases the risk of coronary heart disease.

Thus, when the most plausible effects of an intervention or exposure are relatively large, they can be easily detected through observational studies.[1] But when the most plausible effect size is between 10 and 50 percent, as is the case with many promising interventions, a small amount of uncontrolled confounding could mean the difference between a 20-percent decreased risk, no effect, or even a 20-percent increased risk. While such modest effects are difficult to detect reliably, they can have tremendous public health impact for a common or serious condition. Reliably detecting modest effects of a treatment, however, can only be done through randomized trials.

If such trials are sufficiently large, they eliminate the residual confounding that cannot be controlled in observational studies, by randomly allocating participants to the exposure of interest. For example, if a randomized trial is conducted to test whether beta-carotene reduces the risk of cancer, some participants would be assigned at random to take beta-carotene supplements, while others would serve as the comparison group by receiving no beta-carotene supplements. Such a strategy eliminates the self-selection of exposure that occurs in observational studies, and the impact of other variables that might be more prevalent among those who choose to eat diets high or low in beta-carotene. The unique strength of randomization is that, if the sample is large enough, the two study groups will usually be comparable with respect to all confounding variables, known and unknown, that might independently be related to risk

[1] For an exposure hypothesized to confer harm rather than benefit (e.g., cigarette smoking), randomized trials cannot be justified, because it would be unethical to assign study participants to such an exposure. In such cases, observational studies remain the only epidemiologic study design available, even when the likely effect of the exposure is modest.

of the disease. Randomized trials thus achieve a degree of control over bias and confounding that is not possible with any other epidemiologic design strategy.

Recognizing that small to moderate treatment effects can be reliably detected only with large samples, some researchers have focused on large and simple trials to answer important medical questions. The size of such trials, which generally involve several thousand participants, and the simplicity of their study protocol and streamlined collection of followup data distinguish these investigations from most randomized trials conducted to date.

PRINCIPLES OF LARGE AND SIMPLE TRIALS

The basic principles of clinical trial methodology must be considered in the design and analysis of any randomized trial, regardless of size. However, the design and conduct of large and simple trials rest on several additional principles and considerations (49):

- the need for *Large sample sizes* in order to reliably detect the most plausible small to moderate effects of particular treatments or to exclude with statistical certainty the possibility of such effects,
- the importance of testing widely *practicable treatments* that could have broad application if demonstrated to be effective,
- the use of *broad entry criteria* to determine eligibility for inclusion in trials,
- the use of *streamlined protocols,* and
- the use of a *clinically important outcome measure* to assess the effects of treatments.

Need for Large Samples Sizes

Through the random assignment of treatment, trials maximize the probability that both known and unknown confounding variables will be distributed equally among the treatment groups. Because this phenomenon works "on average," equal distribution is more likely to occur if the trials are large. Moreover, large samples also enhance the statistical power of trials—i.e., the likelihood that a trial will detect an effect if one is truly present.

A fundamental aim of any randomized trial should be to assemble a sample size that is adequate to permit the researchers to definitively detect an effect if it exists, or to clearly demonstrate the lack of an effect if there isn't one. Many randomized trials have failed to provide definitive tests of research hypotheses simply because they were too small to rule out the play of chance as a plausible alternative explanation for any findings that emerged. Such trials can actually do scientific harm if their results are interpreted as providing clear evidence of no effects when the trials simply had inadequate statistical power to answer the research questions with certainty. Null findings have emerged from a number of small trials testing treatments that were later shown unequivocally in investigations with adequate samples to confer clear net benefits.

Two examples of the importance of large samples to definitively evaluate a hypothesis involve the testing of promising treatments for acute heart attacks, or myocardial infarction (MI). The International Studies of Infarct Survival (ISIS) is a set of studies on treatment for MI, conducted through a worldwide collaboration of hospitals, that began in the early 1980s. The first ISIS trial was designed to test the effects of the beta-blocker drug atenolol. More than 16,000 patients in the acute phase of a suspected heart attack were enrolled into ISIS- 1 and assigned at random to receive atenolol (5 to 10 mg intravenously and then 100 mg per day orally for 7 days) or to serve as controls (32). Another trial of beta-blocker therapy, the Metoprolol in Acute Myocardial Infarction (MIAMI) trial, enrolled approximately 6,000 patients to test this treatment (39).

When the two trials were completed, the estimates of the effects of treatment were very similar, with the study participants who received beta-blocker therapy experiencing reductions in vascular mortality of approximately 13 percent in the MIAMI trial and 15 percent in ISIS-1. Though the estimates of effect in the two trials were virtually identical, the ISIS- 1 result achieved statistical sig-

nificance, whereas the MIAMI result did not. This difference in the strength of the conclusions that could be drawn from the two trials resulted almost wholly from their respective sample sizes.

Another promising area of research in the treatment of acute MI in the early 1980s was the use of thrombolytic drugs, agents given during the acute phase of a heart attack to dissolve the clots in the coronary artery that had precipitated the attack. By restoring blood flow to areas of the heart muscle that have been starved of oxygen-rich blood by the blockage, these drugs can spare the heart from permanent damage.

By the mid- 1980s, 24 separate trials had tested the hypothesis that the use of an intravenous thrombolytic agent (primarily streptokinase) would decrease the risk of mortality in patients with acute MI. Of these trials, five reported a statistically significant benefit on mortality from use of a thrombolytic drug, 11 suggested a benefit but were not statistically significant, and eight reported a harmful trend but were also not statistically significant (50). The discrepancies in the trials' findings most likely derived from the fact that the effect of such agents was anticipated to be modest (on the order of a 10- to 30-percent decrease in mortality), and the majority of the individual trials were simply too small to detect such a benefit accurately (none enrolled more than 750 patients).

The uncertainties left by these trials led directly to ISIS-2, in which more than 17,000 patients were randomized to the thrombolytic drug streptokinase or placebo as well as to a month-long regimen of daily low-dose aspirin or placebo (33). With respect to vascular mortality, patients who received streptokinase experienced a statistically significant 25-percent reduction in risk, those receiving aspirin experienced a statistically significant 23-percent decrease, and those who received both treatments experienced a significant 42-percent decrease in vascular death. Thus, this large and simple trial was able to detect definitively the modest but clinically meaningful benefits of thrombolytic therapy in the treatment of acute MI.

The reason that larger trials are better able toreliably detect modest treatment effects derives not just from the numbers of randomized participants but rather from the number of events they experience. For example, whereas a trial of aspirin in the primary prevention of heart disease might require a sample of 22,000 men over the age of 40 in order to detect a 20-percent reduction in risk, a sample of 40,000 women over the age of 45 would be required to detect the same effect, because women have a lower baseline rate of heart disease than men do. Thus, trials must be large enough to accrue sufficient numbers of outcome events to demonstrate either definitive positive results or truly informative null findings.

The identification of effective treatments for a condition also affects the sample size requirements of future investigations. As the efficacy of thrombolysis and aspirin has been demonstrated in large trials, these therapies have become more common components of the routine management of MI patients (35,40). Any new therapies, then, must be shown to confer additional benefits beyond those of an expanding regimen of effective standard treatments. As a result, the absolute magnitudes of any further benefits are likely to be progressively smaller. Such benefits may be very worthwhile, since MI is a common and serious condition, but detecting them will become increasingly difficult and will require trials with even larger samples.

A second circumstance that affects a trial's sample size requirements is the need to compare directly two or more treatments to determine whether one has clear advantages. It was just such a question that led to ISIS-3 (34). Randomized trials had suggested that, in addition to streptokinase, two other thrombolytic agents—tPA (tissue plasminogen activator) and APSAC (anisoylated plasminogen-streptokinase activator complex)—were effective in dissolving clots in acute MI and reducing subsequent mortality. Although thrombolytic therapy was clearly a valuable treatment, it was unclear whether there were any important differences in the benefits and risks of the three prin-

cipal thrombolytic drugs, so a head-to-head comparison of the agents was carried out.

All patients in ISIS-3 received thrombolytic drugs, with one-third of the study participants randomly assigned to each agent. To detect meaningful differences among the treatments, all of which were expected to confer roughly comparable benefits, 1S1S-3 randomized more than 41,000 patients. The trial provided statistically conclusive evidence that there were no significant differences between the three thrombolytic drugs in reducing mortality following acute MI. Moreover, in terms of the most serious adverse effects associated with thrombolytic drugs, tPA and APSAC were shown in ISIS-3 to be associated with significantly more cerebral hemorrhages than streptokinase. The three drugs differ substantially in cost, which ranges from roughly $300 per dose for streptokinase to approximately $1,700 for APSAC and $2,200 for tPA.

A subsequently reported trial, GUSTO (Global Utilization of Streptokinase and Tissue Plasminogen Activator for Occluded Coronary Arteries), which included comparisons of streptokinase and tPA, suggested that a newer method of administering tPA very rapidly conferred a slight advantage in reducing mortality over streptokinase (26). Again, however, tPA was associated with a higher rate of cerebral hemorrhage. Considerable controversy has surrounded specific issues in the interpretation of the GUSTO findings (41).

An issue raised by the findings from GUSTO and other trials of thrombolytic agents is the need to distinguish between differences in treatment effects that are statistically significant, based on comparisons of tens of thousands of patients, and those that are clinically meaningful in treating patients. In the case of streptokinase and tPA, currently available evidence from large-scale trials suggests that emphasizing the differences in thrombolytic agents' efficacy and safety is far less important than encouraging their wider use, since all of them confer clear benefits in a large proportion of acute MI patients (41).

I Testing Widely Practicable Treatments

The need to test widely practicable treatments is another principle of large and simple trials. From a public policy standpoint, a treatment is likely to have a greater effect on public health if it can be readily administered at most community hospitals than if it is very complicated or expensive (or requires specialized training or resources available only at tertiary care facilities), even if the two treatments confer the same degree of benefit.

For example, three recent small randomized trials of treatments for acute MI patients compared the effects of a clot-dissolving thrombolytic agent with those of coronary angioplasty, a procedure in which a balloon-tipped catheter is guided into the blocked coronary artery and briefly inflated to reopen the occluded vessel (22,23,53). In two of the three trials, patients receiving angioplasty experienced lower rates of mortal it y or recurrent MI than did those receiving thrombolytic therapy (23,53). The third trial found no clear evidence of a difference in the effects of the two treatment strategies (22).

These results suggest that the two approaches may be equally effective, or perhaps even that angioplasty has a short-term advantage. Of far more significance from a public health perspective, however, is the fact that only 18 percent of U.S. hospitals are capable of performing angioplasty, with even fewer equipped to conduct emergency coronary bypass surgery (which is necessary in the small number of cases where a vessel abruptly closes following angioplasty). Many acute MI patients in the United States probably live reasonably near hospitals equipped to perform angioplasties as well as emergency coronary bypass surgery, but the widespread use of angioplasty instead of thrombolytic therapy would greatly increase the demands on such facilities and would have tremendous implications for the level of coronary care services required in U.S. hospitals. Consequently, the editorial accompanying the three trial reports concluded that "the strategy of immediate angioplasty for acute myocardial in-

farction has limited applicability because of the severely restricted accessibility of the procedure" (37). At present, therefore, thrombolytic therapy, which can be administered at most emergency care facilities—and in prehospital settings in some areas-can have a far greater overall public health impact on mortality following acute heart attack.

I Use of Broad Entry Criteria

Many randomized trials have studied relatively homogeneous, narrowly defined groups of patients, thereby meeting the scientific urge for precision in knowing exactly which types of patients will benefit from particular interventions. By contrast, large, simple trials generally have used very broad and flexible entry criteria. On a practical level, the use of broad entry criteria aids the recruitment of large numbers of patients and minimizes costs by eliminating the need for elaborate screening procedures. In addition, however, there is a compelling scientific rationale for such a practice (52).

The goal of randomized trials is to provide reliable evidence of treatment effects that can be used to improve clinical practice. Large, simple trials have used very wide entry criteria so that the heterogeneous population under study will more closely mirror the broad population of patients to whom the results can be generalized. A basic premise underlying the use of broad eligibility criteria is that the direction, though not necessarily the magnitude, of the net effect of a treatment is likely to be similar for many subcategories of patients. In other words, the magnitude of any benefit or harm may well differ according to certain patient characteristics, but such quantitative differences in the size of the effect are much more likely than unanticipated qualitative differences, in which one group of participants benefits from a treatment while another either does not benefit or is harmed.

The use of narrow eligibility criteria can unnecessarily limit the generalizability of findings. For example, from the results of animal experiments, researchers thought that thrombolytic therapy would be ineffective or perhaps even harmful if initiated more than six hours after the onset of symptoms. Some early trials of thrombolytic therapy, therefore, restricted participation to patients with symptoms of less than six hours' duration. The rationale for this limitation was that little benefit would accrue to patients whose symptoms were of longer duration but that the drugs' known risks (e.g., cerebral hemorrhage) would still exist. However, even if the results of such trials suggested a benefit of thrombolysis, they could not answer whether the treatment might also benefit patients who arrived at hospitals more than six hours after their symptoms began.

ISIS-2 adopted much wider eligibility criteria, enrolling patients up to 24 hours after the onset of MI symptoms. Large-scale trials can, and indeed should, collect data on key variables that may define clinically important subcategories of patients in whom treatment effects may substantially differ. Therefore, the time that had elapsed since the onset of symptoms was one of the select variables in ISIS-2 for which information was gathered at baseline. The collection of such data allowed for the analysis of trial results according to duration of symptoms prior to treatment, an analysis that demonstrated that the benefit of streptokinase, although greatest for patients treated early, extends to those treated up to 24 hours after the onset of symptoms. Overall, a 25-percent reduction in cardiovascular death was associated with streptokinase treatment given within 24 hours of the onset of symptoms. The reduction was 35 percent for those treated within four hours and 17 percent for those treated within five to 24 hours.

Although the ISIS-2 results demonstrated the advantages of wide eligibility criteria, precise definition of the eligibility criteria for a trial is a matter of scientific judgment, based on the particular question being asked. Randomizing patients up to one week following the onset of MI symptoms, for example, makes little biological sense, in view of the known properties of thrombolytic drugs and the pathophysiology of MI over such a period. Not only would such broad eligibility criteria unnecessarily expose late-treated patients to the possible risks of thrombolysis, but they would also dilute

any benefit of treatment to such an extent that the overall finding from the trial might be null, even if analyses restricted to early treated patients suggested a clear benefit. In fact, while some early trials of thrombolysis used unduly restrictive entry criteria, others cast too wide a net, randomizing patients up to 72 hours after initial symptoms (50). Thus, reasonable judgments must be made not only in identifying the population at risk for the outcome under study, but in defining the group of individuals in whom an effect of the intervention is biologically plausible.

One criticism of the use of broad entry criteria is that even though the study's overall results may apply to a wide population of patients with a particular disease, they do not offer much guidance about how to treat individual patients with specific medical profiles. This tension—between the broadly relevant data available from large, simple trials and the highly detailed information upon which practicing clinicians might ideally wish to base individual treatment recommendations—may never be fully resolved. However, several factors support the use of wide, rather than narrow, entry criteria in many trials evaluating promising medical interventions. First is the belief that unless there is a clear reason to believe otherwise, a beneficial treatment is likely to be effective across a broad spectrum of patients. Results from trials using broad entry criteria, therefore, are directly relevant to the wide spectrum of patients to whom the results will be generalized in actual clinical medicine. Second, if the effect of an intervention differs among categories of patients, a large-scale trial enrolling a broad range of patients might be the only way to detect the differences. Even in large trials, however, the statistical power to detect treatment effects among subcategories of patients may be inadequate. Further, if many subcategories are analyzed, it becomes increasingly likely that an erroneous finding will emerge simply from the play of chance. Therefore, for research questions that require the enrollment of large numbers of participants, the main finding will be one that answers whether, on average, the study intervention confers a net benefit compared with no treatment (or the alternative treatment). More precise evidence may emerge from analyses of select subcategories, but applying trial results to medical practice will always involve making individual clinical judgments based on each patient's medical profile. A decade ago, a paper describing the principles of large, simple trials addressed these issues succinctly:

> Trials are at least a practical way of making some solid progress, and it would be unfortunate if desire for the perfect (i.e., knowledge of exactly who will benefit from treatment) were to become the enemy of the possible (i.e., knowledge of the direction and approximate size of the effects of the treatment of wide categories of patient) (49).

Use of Streamlined Protocols

The use of streamlined study protocols has very practical advantages in the design of a large-scale trial. If a trial requires many thousands of patients in order to answer a question reliably, the trial organizers usually must reach beyond the confines of the academic medical centers (where most research is conducted) to involve general-care community hospitals or even medical settings in a number of countries. This can be accomplished only if treatments can be administered in a wide range of settings, as is the case for thrombolytic therapy. Furthermore, to secure the cooperation of busy physicians and nurses (whose primary mission is to care for their patients, not to conduct research), trial treatments must be relatively simple to administer, and the added burdens of participation must be minimized whenever possible by using streamlined screening procedures and collecting only the most important followup data needed for assessing the efficacy and side effects of the treatment.

The cost of research is also an important factor in the move toward simple trial protocols. Particularly during an era of shrinking research budgets and increased competition for funding, efficient study designs are imperative if large trials are to be funded to any significant extent. For example, the Beta-Blocker Heart Attack Trial (BHAT), which began in 1977, randomized 3,837 patients with prior heart attacks in order to test whether the beta-

blocker drug propranolol hydrochloride reduced total mortality, at a total cost of $20 million (3). In contrast, a trial testing the drug digitalis among patients with congestive heart failure, which was begun in 1991 and is employing a streamlined trial protocol, randomized 7,790 patients and will have a total budget of $16 million (21). After adjustment for inflation, the earlier BHAT investigation cost approximately $11,350 per participant, while the ongoing digitalis trial will incur costs of approximately $2,050 per participant.

Similar efficiencies are possible in studies of preventive interventions in apparently healthy participants. Most such investigations have collected extensive baseline and followup data and required regular clinic visits, with costs generally ranging from $3,000 to $15,000 per randomized participant for a five-year trial. In contrast, the Physicians' Health Study, a trial testing aspirin and beta-carotene in the prevention of cardiovascular disease and cancer, has been conducted entirely by mail among 22,071 U.S. male physicians at a cost of approximately $80 per participant per year (4).

Practical considerations underscore the need for simple trial protocols, but in addition, the most widely practicable treatments are often those that are simple. And for interventions where the outcome of interest is a straightforward, easily ascertainable event such as mortality, most of the crucial information needed for future clinical decisionmaking and public health policy is available from the streamlined data collected in large, simple trials.

In ISIS-2, for example, virtually every patient entering a participating hospital within 24 hours of the onset of symptoms of suspected MI was considered eligible to participate. If there were no clear indication for or against the trial treatments, the patient was eligible to be randomized. If informed consent were obtained, a 24-hour toll-free randomization telephone line was dialed, and the physician or nurse collaborator provided basic identifying data on the patient as well as information about a very few select medical variables, such as time since the onset of symptoms. A randomization code was then obtained and matched against one of the treatment packs stored in the hospital, and the contents of the pack administered to the patient. At the time of the patient's hospital discharge, the clinician completed a simple one-page followup form, providing information on vital status (i.e., whether the patient was alive or dead) as well as major in-hospital events, such as reinfarction, stroke, or significant bleeding episodes. The clinician then sent this form, along with the results from a pre-randomization electrocardiogram, to the international coordinating center in England. At that point, the clinician responsibilities to the trial were over.

An important assumption underlying the use of a simple protocol with streamlined followup is that the areas of chief concern regarding adverse effects of the intervention have been reliably identified. Although the balance between the benefits and the risks of a treatment is unknown-and, indeed, is the principal question being asked in most large trials—preliminary testing or knowledge of biological mechanisms should have allowed the researchers to identify the most serious potential side effects so that the collection of followup data could be confined to a few key variables. Trials of agents or procedures for which there is little prior knowledge concerning safety may require much more detailed data collection and thus will more closely resemble traditional randomized controlled trials.

I Use of Clinically Important Outcome Measures

Small and more complex trials may be important early in the development of a treatment. Such investigations may collect data on scores of variables to assess their response to treatment. This may, in turn, provide important information about the action of the drug, its side effects, or features of the disease itself. When an intervention is sufficiently promising to warrant testing for efficacy in a large-scale trial, however, the fundamental goal is to obtain information that can inform clinical practice and public health policy. For this reason, the primary outcome in a large, simple trial should be a clinically meaningful event, not an intermedi-

ate marker whose clinical significance is unknown. In most trials of serious diseases, the fundamental question is whether a treatment increases patients' chances of survival. Major morbidity events, such as nonfatal heart attacks, may also be suitable endpoints in some trials, but the use of subclinical or intermediate markers as surrogates for clinical endpoints can lead to spurious conclusions.

Reliance on an intermediate endpoint in studies of the effect of thrombolytic drugs in the treatment of acute MI, for example, may have led to erroneous conclusions about the relative benefit of different agents. Many physicians believed that the thrombolytic drug tPA was superior to streptokinase because it appeared to be faster at dissolving the clots in the coronary artery that precipitated the attack. This conclusion was based on angiographic studies demonstrating that 90 minutes after treatment, blood flow was restored through the previously occluded artery in 70 percent of patients receiving tPA compared with 50 to 55 percent of patients receiving streptokinase (46). However, further studies indicated that coronary patency rates for tPA and streptokinase become equal over the next several hours. Moreover, for the primary clinical endpoint of mortality, the results of large-scale trials demonstrated identical 35-day vascular mortality rates for patients given tPA and those given streptokinase (25,34).

In addition to making a clinically important outcome the primary focus, a large and simple trial must also have a main outcome event that can be fairly readily ascertained without extensive, specialized testing or frequent in-person followup visits. In this regard, mortality is the most straightforward outcome event, inasmuch as its occurrence is not subject to dispute and can even be tracked by searching death certificate databases or using other indirect methods of followup. Nonfatal medical events may also be suitable endpoints for large, simple trials. For example, most nonfatal heart attacks or cancer diagnoses can be verified using existing medical record information that would be available regardless of whether an individual was part of a trial protocol.

OTHER APPLICATIONS OF LARGE, SIMPLE TRIALS

Acute MI has been the clinical context in which the principles of large, simple trials have been most widely applied to date, as discussed above. Because it is an easily defined, common, and serious clinical event—and one for which the fundamental measure of a treatment's efficacy can be made over a relatively short time frame—acute MI is particularly well-suited to this research approach. In addition, however, trials employing these principles have been conducted and proposed for a wide range of treatments and health conditions, including longer-term trials of chronic heart disease, the management of women with high-risk pregnancies, treatments for patients with human immunodeficiency virus (HIV) infection or acquired immune deficiency syndrome (AIDS), and the surgical treatment of cancer, as well as the testing of promising interventions in the primary prevention of cancer and heart disease among apparently healthy participants.

Polio Vaccine Field Trial: An Early Example

Perhaps the first large and simple randomized trial was carried out 40 years ago, when the National Foundation for Infantile Paralysis recruited a team of physicians and public health researchers to mount a massive randomized trial to test the efficacy of the Salk polio vaccine (38). More than 400,000 U.S. school children took part in this experiment in the spring and summer of 1954. The polio vaccine trial randomly assigned half of the participants to receive the vaccine, while half received a placebo injection. The incidence of disease in the two groups was then tracked by simply monitoring the hospitalizations for polio in the areas where the field trial was carried out. Over the course of several months, the effectiveness of the vaccine in preventing this serious, disabling childhood disease became clear (20).

In many respects, the large and simple design of the massive polio trial was a response to the urgency of the problem, in which there was tremendous

pressure to provide a quick and reliable test of the newly developed vaccine in a single polio season. While the polio trial may have been the first example of a large and simple randomized trial, it was not until several decades later—in the late 1970s and early 1980s—that the principles of this approach to answering health questions were more formally described and its methods more widely used to evaluate clinical questions (38).

Digitalis in the Treatment of Congestive Heart Failure

Researchers have recently begun testing the drug digitalis in the treatment of congestive heart failure in a long-term trial that has incorporated many features of large, simple trials (12). Although overall rates of cardiovascular disease have declined significantly in the United States, over the past two decades the incidence and prevalence of congestive heart failure (CHF) have increased significantly, a pattern that is expected to continue as the population ages. CHF, a cardiac syndrome characterized by a weakening of the contractions of the heart muscle, is estimated to be a primary or contributing cause of 250,000 deaths in the United States each year.

Digitalis preparations, which have been available for more than 200 years, are one of the most commonly prescribed treatments for CHF. In 1986, more than 12 million prescriptions for this drug were written in the United States (12). Despite its widespread use, the net effect of this drug on mortality in patients with CHF remains uncertain. Although a number of small trials of digitalis have been conducted, the results of the trials are inconsistent (51). Digitalis may improve the output of blood by the heart (ejection fraction) and thereby slow the progression of CHF and decrease mortality, but the drug has other biochemical properties that, in theory, may increase the risk of dangerous changes in cardiac rhythm. In view of the continued uncertainty regarding the net effect of digitalis, a large trial of the drug was initiated in 1991 under the direction of the National Heart, Lung and Blood Institute (NHLBI).

Because any large benefit of digitalis would probably have been clear from the smaller studies already conducted to date, its true benefit-if any-in reducing mortality is likely to be on the order of 10 to 15 percent. However, as with thrombolysis or aspirin therapy of acute MI, even a modest mortality benefit for such a common condition could be of great public health value. For researchers to detect a benefit of digitalis in this range, about 2,000 deaths would need to occur in a trial population. To enable researchers to observe this number of events over a relatively short period, the trial has enrolled nearly 8,000 patients at more than 300 hospitals throughout the United States and Canada.

The digitalis trial will involve treatment and followup of patients for three years. The principal entry criterion for the trial will be moderate or severe CHF (ejection fraction< 0.45). All patients in the trial must have had a chest x-ray within the past six months and cardiac ejection fraction documented by either angiography or echocardiogram. Patients will be randomized via telephone calls to a central coordinating center, with key baseline data given directly by phone for entry in the study's database. Each randomized patient must return for a followup visit in four weeks, and every four months thereafter. Because digitalis has a relatively narrow therapeutic window, with high toxicity at elevated doses, blood will be drawn during followup visits to monitor the serum levels of digitalis as well as those of potassium, creatinine and magnesium. In addition, since the appropriate dose of digitalis depends on patient characteristics such as age, weight, and sex, four different dose regimens will be used.

Despite these considerations, which add complexity to the trial protocol, the digitalis trial retains two chief characteristics of large, simple trials: the collection of followup data is limited (a one-page questionnaire at each visit), and the clinic visits and many of the monitoring tests of dose level, as well as the required radiologic studies for eligibility, would be carried out anyway as part of the standard clinical management of CHF pa-

tients. The principal outcome measurement in the study will be mortality, and the trial should provide clear evidence of digitalis's net effect on mortality in patients with CHF.

Aspirin in Treatment of High-Risk Pregnant Women

Pre-eclampsia, a condition caused by high blood pressure, is a common and serious complication of the second half of pregnancy. It can lead to intrauterine growth retardation and fetal death as well as to complications of prematurity, because early delivery of the baby is the only effective approach to the condition.

Several trials of aspirin have suggested that treatment in high-risk pregnant women is beneficial, but the small samples of patients in most of the trials have left a great deal of uncertainty concerning the treatment's effects. To address this problem, the Collaborative Low-Dose Aspirin Study in Pregnancy (CLASP) randomized 9,364 women in 16 countries to either 60 mg of aspirin or a placebo daily (8). According to the trial's broad entry criteria, women were eligible if they were between their 12th and 32nd weeks of pregnancy and were judged by their treating clinicians to be at sufficient risk of pre-eclampsia to consider aspirin treatment. Randomization was carried out by having clinic staff telephone a 24-hour randomization service. For each patient, data were collected on several key variables at entry, and a single-page followup form was completed following hospital discharge at the end of the pregnancy, recording information on treatment compliance, use of other drugs, and major clinical events occurring after randomization.

Overall, those assigned to receive aspirin experienced a 12-percent reduction in the development of pre-eclampsia, but this difference was not statistically significant. Aspirin-allocated women did experience a modest but significant lower rate of delivery before 37 weeks estimated gestation. However, there were no significant differences between treatment groups in the proportion of stillbirths, neonatal deaths, or babies with intrauterine growth retardation. Because of the possibility that the benefits of aspirin might be restricted to certain subgroups of women, the CLASP protocol called for separate analyses of data based on several entry characteristics. There were no subgroups in whom the reduction in pre-eclampsia was as large as that reported in the earlier small trials. The authors concluded that currently available data do not support the widespread use of aspirin in women at high risk for pre-eclampsia. Nonetheless, among women with preterm deliveries, there was a significant trend toward greater reductions in the development of pre-eclampsia in the group that received aspirin. The authors suggest that aspirin may have effects in women who are susceptible to early pre-eclarnpsia that it does not have among women who develop this condition in the late stages of gestation. Aspirin may, therefore, be justified in those at particularly high risk of early-onset (before 32 weeks) pre-eclampsia, but inasmuch as these women are difficult to identify prospectively, the clinical implications of the CLASP findings may be restricted to high-risk women with prior histories of early-onset pre-eclampsia.

Treatments for Patients with HIV or AIDS

Several investigators have suggested that large, simple trials could be used for the efficient testing of potential treatments for those infected with HIV or with diagnosed AIDS (6,7,13,14,42). Very detailed studies in specialized centers are clearly crucial to gain more knowledge about this disease. Indeed, it is from the intensive study of patients and potential treatments that promising hypotheses will emerge. To reliably answer the broader question of a treatment's net clinical effect, however, will require collaborative trials using the principles of large, simple trials, because most of the promising therapies are likely to have only small to moderate effects. In addition, as with the treatment of acute MI, trials will need to be designed to detect the equal or superior efficacy of new treatments in relation to an expanding array of standard therapies.

Both the National Institutes of Health and the American Foundation for AIDS Research have es-

tablished networks of community physicians for research studies. Such consortia could form the organizational basis for the implementation of large, simple trial protocols (13, 14). Another avenue that has been suggested for the development of large trials is the enrollment of patients now receiving treatments as part of the system known as the treatment IND (investigational new drug) or *parallel track*. This expanded-access program was approved by the Food and Drug Administration to make treatments still undergoing experimental evaluation available to a broad population of patients who have life-threatening diseases and who are no longer able to tolerate or benefit from the standard available treatments. Although these programs are providing many patients with experimental AIDS drugs and uncontrolled observational followup, direct comparison of such treatments could be carried out as part of simple, randomized treatment protocols, without undue requirements for additional work by busy clinicians, but with systematic coordination by a data center.

Because large-scale community-based trials would collect uniform data on only a small number of important variables, more extensive data could be gathered at selected participating sites, such as academic research centers. This strategy, which has been used in other large trials, may be particularly appropriate for AIDS treatment, where the rapid development of new experimental therapies means that there is frequently much less long-term experience with a drug's toxicity orother effects than is often the case with agents being tested in large-scale trials. Such trials might, therefore, more appropriately be considered hybrid trials, with a large component that uses a simple trial protocol and a small subgroup for whom a more detailed randomized clinical trial protocol is implemented.

Since a major goal of AIDS treatments is to prolong survival, large numbers of patients must be enrolled in trials if any net benefit of these treatments on mortality is to become known relatively quickly. New antiretroviral agents, for example, could be compared with current standard therapies in large-scale trials to assess their survival benefits (42). Other important questions that could be answered through large trials include determining the optimal doses of available treatments (13). Many currently available AIDS treatments have significant toxicity. Randomized trials comparing different doses of a particular drug could determine whether lower, less toxic doses confer a similar survival benefit. Studies of zidovudine (AZT), for example, have already shown that daily doses of 600 mg are as effective at prolonging survival as 1200-mg doses. Even lower doses might work equally well, and such a finding could significantly improve the quality of life for many AIDS patients (13).

At present, AIDS treatments differ qualitatively from those used in other conditions. In the case of acute MI, where treatment with thrombolysis or aspirin saves several lives for every 100 patients treated, these individuals are, in some sense, considered "cured" because they avoided death during the high-risk period immediately following their attack. Such patients remain at higher risk for cardiovascular death, but they could live for decades and then die from nonvascular causes. There is no comparable life-saving effect of current AIDS treatments, which confer only short-term survival benefits.

This suggests some justification for using other clinical endpoints besides mortality, such as quality-of-life measures or the development of opportunistic infections, to determine the benefits of some agents. It is crucial to keep in mind, however, that an observed improvement in such outcomes may not translate into longer patient survival. The demonstrated benefits of any treatments approved on such a basis must be clearly identified to avoid overstating their known effects.

Biological markers, such as CD4 cell counts, have also been proposed for use as endpoints in AIDS trials. Because drugs may affect these surrogate endpoints much sooner than clinical outcomes (e.g., opportunistic infections or death), the use of such endpoint markers can reduce the needed size and duration of a trial. This approach is clearly attractive in the face of a fatal epidemic. Unfortunately, none of the biological markers currently

measured in AIDS patients has been shown to predict clinical course or survival reliably enough for use as a firm endpoint (7).

After the initial demonstration that AZT confers a short-term reduction in the mortality of symptomatic patients (19), a trial was conducted to test whether early treatment would delay the onset of AIDS in asymptomatic individuals infected with HIV. A clear delay in disease onset was observed and the trial was stopped prematurely (47). However, it was not at all clear whether the early use of AZT in asymptomatic patients would extend their survival beyond what would be achieved by initiating therapy at the onset of the disease. Moreover, early use of an antiretroviral agent may render the drug less effective later, during the actual disease phase, thereby potentially shortening the survival time after the development of full-blown AIDS (42). Because of the debilitating and fatal nature of the disease, this may be an acceptable choice to patients, who face limited life expectancies regardless of which treatment course they pursue. Information on this question should be available, however, so that patients can make informed choices.

To provide further data on the relative merits of immediate versus delayed treatment with AZT, a randomized controlled trial in Europe compared how the two treatment approaches affect mortality (1,9). The Concorde trial was a multicentered trial carried out in England, Ireland, and France among 1,749 HIV-infected individuals who were symptom-free at baseline. Half of the participants were randomized to begin immediate treatment with AZT; the others were randomized to deferred treatment, which entailed taking inert placebo pills that resembled AZT. Once patients exhibited symptoms of AIDS or AIDS-related complex (ARC), or had persistently low CD4 cell counts that led their physicians to believe treatment was indicated, their assignments were unblinded and those who were receiving placebos began AZT therapy.

Throughout the trial, the patients randomized to immediate AZT treatment had significantly higher CD4 cell counts. It has been postulated that higher levels of these disease-fighting cells indicate the efficacy of an AIDS treatment, and that decreases in CD4 cell counts signal the progression of HIV disease. Despite the favorable effect of immediate treatment on CD4 cell counts in asymptomatic patients, the three-year survival rates in the two treatment groups were virtually identical (92 percent in the immediate treatment group vs. 94 percent in the deferred therapy group). Even more surprising, the Concorde results indicated that early treatment of HIV did not appear to slow the rate of progression of asymptomatic HIV disease to ARC, AIDS, or death—a finding in marked contrast to previous studies, which had indicated a benefit of early treatment of asymptomatic patients. However, these tria~s were stopped much earlier than the Concorde trial. Short-term followup data from the Concorde trial were also compatible with the finding of a benefit from early treatment, but the apparent advantage of immediate AZT therapy disappeared with longer-term treatment and followup.

Although the Concorde findings appear to rule out any large benefit from early treatment with AZT in asymptomatic individuals, the trial was not large enough to rule out the possibility of a small advantage of such treatment.

The Concorde results raise important questions about the ultimate public health benefit of the rapid approval of AIDS drugs in the United States. The highly organized activities of individuals with HIV mark an unprecedented degree of direct involvement by affected patients in the quest for advances in treatment of their condition. This activism has led to many positive changes in what some have regarded as an often cumbersome and unduly bureaucratic drug approval process. At the same time, however, the pressure to speed drug approval may also lead to rapid decisions made without full benefit of the optimal quality or quantity of randomized trial data.

The guiding principle of broad entry criteria in large, simple trials has particular relevance to the study of AIDS treatments. Many AIDS patients are interested in participating in treatment protocols, but have been excluded because of stringent entry criteria. Because most of the treatments that are found to be effective will be made available to

most patients, it is reasonable—and indeed desirable—to include a broad range of patients in trials (6).

As in the large trials of patients with MI or CHF, several key baseline variables should be collected to allow for the assessment of any differing effects of treatments among subgroups.

Breast Cancer Treatments

Some of the most significant advances in the treatment of breast cancer have resulted from the collaboration of a large number of hospitals in the National Surgical Adjuvant Breast Project (NSABP), which has coordinated multicentered randomized trials comparing different treatment approaches to breast cancer (17, 18). For many decades, the standard treatment was radical mastectomy, which entails removal of the breast, axillary lymph nodes, and pectoral muscles. Anecdotal evidence suggested that less disfiguring approaches might be as effective as more extensive surgery, but definitive evidence was not available to settle the debate.

In a study to determine whether alternative treatments to radical mastectomy increased the risk of cancer recurrence or death (18), a total of 34 institutions in the United States and Canada randomized 1,665 women with operable breast cancer. Women judged to be free of cancer in axillary nodes were randomly assigned to undergo radical mastectomy, total mastectomy with regional radiation treatments, or total mastectomy alone. Those judged to have cancer in the axillary nodes were assigned randomly to either radical mastectomy or total mastectomy with radiation treatment. The overall rates of survival and cancer recurrence were similar for all three groups of patients with clinically negative axillary nodes. The overall survival at 10 years for patients with positive axillary nodes was similar for those who underwent radical mastectomy and those who had total mastectomy with accompanying radiation treatment. This trial provided clear evidence that surgery less extensive than radical mastectomy could be safely performed with no decrease in long-term survival.

A second trial coordinated by the NSABP sought to determine whether even greater breast conservation could be safely achieved through segmental mastectomy (also referred to as lumpectomy), in which only the tumor and immediately surrounding tissue are removed (17). This trial randomized 1,843 women who had breast tumors no more than 4 cm at the largest dimension. The three types of treatment tested were: total mastectomy, segmental mastectomy (lumpectomy), or segmental mastectomy with accompanying radiation treatments. All of the women underwent removal of their axillary nodes, and those patients found to have evidence of nodal cancer underwent chemotherapy. After five years, the overall rates of survival were better for the women who had received segmental mastectomy, with or without radiation, than for those who had undergone total mastectomy, and the rates of survival with no recurrence of the disease recurrence were better for those who had undergone segmental mastectomy with radiation treatment than for those who had undergone total mastectomy.[2]

Because the NSABP trials enrolled a broad range of patients, their results are clinically relevant to a large proportion of women, and the outcomes measures-disease-free survival and overall survival-were easily ascertained, clinically important events. The trial treatments and follow-up monitoring were clearly more complex than those of nonsurgical large-scale trials, but most of the procedures are part of the standard manage-

[2] The NSABP has made enormous contributions to the field of breast cancer treatment, which would not have been possible without such large-scale collaborative trials. Recently reported evidence of misconduct by one of the NSABP clinical collaborators has raised concern about the scientific integrity of the findings concerning segmental mastectomy. However, the principal findings were unchanged in a re-analysis that was performed excluding data on patients who may have been inappropriately entered into the trial (16).

ment of breast cancer patients and would have been followed anyway.

Promising Therapies in Primary Prevention

For a common and serious condition such as acute MI, even modest reductions *in* mortality can have a significant public health impact, saving tens of thousands of lives per year. At the same time, effective means of preventing such a disease could, in theory, have a far greater impact, preventing perhaps hundreds of thousands of deaths each year. The conduct of large-scale trials of promising therapies in primary prevention presents unique challenges not faced by those conducting trials to test treatments in a population with a specific disease.

One primary prevention trial that has employed many of the principles of large and simple trials is the Physicians' Health Study, an ongoing, randomized, double-blind, placebo-controlled trial begun in 1982 to test the effect of low-dose aspirin on cardiovascular disease and beta-carotene on cancer risk among a population of 22,071 apparently healthy U.S. male physicians, ages 40 to 84 (28,29).

Because the rates of disease and death among an apparently healthy population "at usual risk" are much lower than rates among a comparable group of individuals with a serious condition such as MI, a primary prevention trial must not only enroll a large number of participants, but also follow them for an extended period in order to allow for valid tests of the study's hypotheses. The participants must also remain compliant with their assigned treatment regimens and be conscientious in maintaining contact with the researchers to report their health experiences. Significant noncompliance with the assigned study regimens or losses to followup will weaken the study's ability to generate valid results. The choice of a study population for such a trial, therefore, is particularly important. Because of their clear interest in health issues, physicians were considered a group who would be motivated participants willing to follow daily pill-taking regimens for an extended period.

A guiding principle of large, simple trials is to minimize the necessity for procedures or clinic visits beyond those that would take place in the standard management of a condition (49). But participants in a primary prevention trial are, by definition, free of major disease and therefore have no regular clinic visits or procedures. Because physicians were deemed capable of reporting on their own health with a high degree of accuracy, the trial could be conducted entirely by mail. Annual supplies of study medications are sent in convenient monthly calendar packs, and brief, followup questionnaires are mailed to collect data on compliance and relevant outcomes at 12-month intervals. Reports of study outcomes are verified by seeking permission to obtain copies of confirmatory medical records.

The Physicians' Health Study also implemented a prerandomization run-in period, in which the participants took their study pills for approximately 18 weeks before their official randomization into the study took place. After this run-in period, the doctors were sent a brief followup questionnaire, and only those who reported that they took their study pills at least two-thirds of the time were randomized into the trial. Participants were randomly assigned to one of four treatments: aspirin alone, beta-carotene alone, both active agents, or both placebos. This efficient design (known as a 2x2 factorial design) has allowed the trial to test two agents simultaneously, with little increased cost over that of a study testing one agent alone. The choice of study population, the enrollment procedures, and the prerandomization run-in period were designed to assemble a group of proven compliers who would be likely to follow the study regimen and report accurately on their health experience for the extended period of trial treatment and followup. Compliance with the assigned regimen is an important factor in determining a trial's statistical power to answer a research question. The run-in, therefore, increased the trial's power. Indeed, a study- population of 22,071 men who remain compliant with the regimen will have greater statistical power to answer a question than would a group of 33,000, one-third of whom become noncompliant (36).

The initial assembly of the study population involved much more effort than that of many large, simple trials, but once the participants were randomized, the trial procedures and followup in the Physicians' Health Study have proved remarkably streamlined and simple. Moreover, compliance rates after 10 years remain at over 80 percent, morbidity followup is over 95 percent, and every death among participants has been recorded.

In 1988, the external Data Monitoring Board terminated the aspirin component of the trial early because the group taking aspirin had experienced a statistically extreme 44 percent reduction in the risk of a first MI. The beta-carotene component of the trial has continued uninterrupted and is scheduled to end in 1995.

In addition to performing added work to assemble its study population, the Physicians' Health Study differed from many other large, simple trials in that it enrolled a relatively homogeneous study population. From the standpoint of generalizability of the study findings, testing these hypotheses in a more heterogeneous group might have seemed preferable. From the standpoint of validity, however, in view of the need to conduct the trial efficiently by mail, to maintain the followup of all participants, and to rely upon their own reporting of health outcomes and motivation to adhere to the treatment regimen over an extended period, a group of health professionals was considered an ideal population in which to conduct the trial. The increased validity and efficiency derived from choosing physicians for the study population were judged, overall, a greater asset to the generalizability of the trial results than a more representative study population unable to maintain adequate compliance for the duration of the study.

Although it may be reasonable to assume that the direction of any net effects seen in the Physicians' Health Study would be similar for other groups, the balance of benefits and risks may well differ for populations with different risk profiles. Because of the desirability of obtaining direct evidence of aspirin's primary prevention effect in women, a large-scale trial of aspirin was begun in 1992 among apparently healthy U.S. female health professionals. The Women's Health Study plans to enroll approximately 40,000 women, ages 45 and older, and will test the effects of low-dose aspirin, beta-carotene, and vitamin Eon the risks of cardiovascular disease and cancer (5).

IMPACT OF LARGE, SIMPLE TRIALS ON CLINICAL PRACTICE

Although the immediate goal of any large-scale randomized trial is to provide a reliable test of the intervention, the ultimate goal is to provide information that can be incorporated into the clinical management of patients in general medical practice. It is difficult to draw broad generalizations about the effects of large, simple trials on medical practice. Nevertheless, because large trials are generally better than small trials at answering research questions, their degree of scientific reliability is high, an important factor that probably influences clinicians' receptivity to research findings. The results of large trials also tend to be published in prominent journals, making physicians and the public more aware of them.

Several reports suggest that the recent large, simple trials of the treatment of acute MI have indeed had a measurable impact on medical practice. For example, in the mid- 1980s, about 25 percent of acute MI patients in the United States received thrombolytic therapy. By 1989, following publication of the ISIS-2 results, this figure rose to just under 40 percent (40). Aspirin was administrated to 39 percent of all acute MI patients before the ISIS-2 report, and approximately 72 percent of all patients following the report (35). Significantly more patients could benefit from thrombolysis and aspirin in acute heart attacks than are presently receiving these treatments (31). Nevertheless, the data indicate that practitioners are adopting these treatment regimens, which have been demonstrated to confer clear benefit.

Although findings from a trial maybe clear and unequivocal, in terms of their public health impact, it is equally important that they address questions considered important by clinicians. For this reason, the ability to provide reliable data on

important possible modifying factors, such as age or time delay to treatment for thrombolysis, may also be a determinant of the impact of the findings for a large, simple trial on medical practice.

Large and simple trials may also affect clinical practice more directly and immediately than small, academic-based investigations because the large trials involve the collaboration of the practicing physicians who ultimately decide how to incorporate research results into patient care (48). Perhaps the best example of this is a series of trials of acute MI treatment in Italy. The GISSI trials (Gruppo Italiano per 10 Studio della Streptochinasi nell'Infarto Miocardico) have tested questions similar to those addressed in the ISIS trials. The first GISSI trial paralleled ISIS-2, testing intravenous streptokinase versus standard treatment among 11,806 patients at 176 coronary care units throughout Italy (24). Three-fourths of Italy's coronary care units collaborated in the trial. Within a year after the publication of the trial finding demonstrating a clear benefit of streptokinase, the drug had become routine treatment for acute MI in 99 percent of the country's coronary care units, suggesting a remarkably rapid and complete incorporation of the trial results into clinical practice (45). The GISSI investigators also held symposia for collaborating clinicians in order to provide scientific background on the trial treatments as well as information about aspects of randomized trial methodology.

LIMITATIONS OF LARGE, SIMPLE TRIALS

Testing a promising intervention by assembling a large study population and using a streamlined trial protocol can be an extremely effective approach to answer a wide range of important health questions. However, there are many questions for which large, simple trials are either not necessary or not feasible.

The anticipated number of outcome events is the primary determinant of how big a sample must be in order for a trial to reliably detect a moderate treatment effect. In a primary prevention trial, the sample must be quite large. In the Physicians' Health Study, where the anticipated rate of cardiovascular disease was very low, more than 22,000 participants were randomized in order to detect meaningful differences between the group receiving aspirin and the group receiving a placebo. Fewer than 400 heart attacks occurred during a five-year period (139 in the aspirin group versus 239 in the placebo group). The 44-percent decrease in the risk of a first MI among those assigned aspirin was highly statistically significant, with probability of less than 1 in 100,000 that the finding was the result of the play of chance (p-value < 0.00001) (44). A trial one-fourth the size of the Physicians' Health Study, with 5,500 participants (which would still be substantially larger than most randomized controlled trials), would have had inadequate statistical power to detect with certainty the 44-percent reduction in mortality.

By contrast, in a trial testing a chemotherapeutic agent inpatients with advanced metastatic cancer, half of whom were expected to die within six months, a small trial of several hundred patients could reliably detect the fact that six-month survival could be achieved in three-quarters rather than half of the patients. Large-scale trials, therefore, are not needed to detect even modest treatment effects in a population where the expected outcome rate is extremely high.

With respect to the trial treatment and study protocol, large, simple trials may not be feasible for studying interventions that are complex to administer and that require frequent clinic visits. Examples include trials testing new physical therapy treatments or drug trials in which frequent blood tests are necessary to maintain proper dosing levels and to monitor potential toxicity.

Also impractical for study in large, simple trials are interventions for which information on intermediate biological markers is deemed important for understanding the treatment's actions or aspects of the pathophysiology of the disease. For example, in a trial of a potential AIDS treatment, detailed laboratory studies at regular intervals may be important, because there maybe comparatively limited knowledge concerning the drug's side effects as well as the postulated mechanism

for its benefit. An alternative to a highly complex investigation may be to carry out more intensive laboratory studies among a subset of participants, while the overall trial design incorporates features of the large, simple trial approach.

Finally, the principle of evaluating a treatment's effect on a clinically important outcome that is easily ascertained also limits the types of questions that can be addressed by large, simple trial designs. Examples in which the outcome of interest may be clinically important but difficult to assess in a streamlined trial include tests of promising treatments for arthritis, where objective measures of mobility must be made; mental illness, where detailed clinical evaluations are needed; visual acuity; and many quality of life outcomes, which are measured in terms of competence in carrying out activities of daily living.

ADDITIONAL ISSUES

Timing of Randomized Trials

Timing is an important issue in the initiation of all randomized trials. For ethical reasons, there must be sufficient belief in the potential benefit of a drug or procedure to justify exposing half the individuals to it, while at the same time, there must be sufficient doubt about its efficacy to justify withholding the intervention from the other half. Ideally, therefore, randomized trials should be conducted as soon as there is a belief that a treatment might confer a net benefit.

The danger in waiting too long to conduct a trial of an intervention that is potentially effective, but as yet unproved, is that it may be adopted into widespread clinical use and become accepted as standard therapy, even without firm evidence supporting its efficacy. Not only does this expose individuals to medical interventions that might not be beneficial (or might even be harmful), but it also increases the logistical difficulty of conducting subsequent randomized trials to evaluate the intervention. Such trials may not be feasible if potential participants or health care providers become reluctant to be part of a trial in which some participants will not receive a treatment that has come to represent standard medical care. It may be difficult to find a sufficiently large population of individuals willing to forego a treatment or practice believed to be beneficial for the duration of the trial, even if there is no sound evidence to support this view.

For example, radical mastectomy for breast cancer gained wide acceptance as the standard of care after its introduction in the early 20th century by William Halsted. Halsted's clinical impression was that removal of the surrounding lymph nodes and muscle, in addition to the breast itself, would decrease the risk of recurrence or spread of the cancer and subsequent mortality. By the 1970s, however, questions that had been raised concerning the necessity of the radical mastectomy (11) prompted randomized trials comparing this procedure to less extensive surgery. Many physicians resisted the call to randomize patients into such a trial, because the radical mastectomy was so entrenched as the standard of care (2).

Eventually, however, a large number of women were enrolled in two multicenter trials. The trials clearly demonstrated that for women with localized tumors and no evidence of spread beyond the breast itself, five-year mortality rates as well as overall rates of recurrence were similar in those undergoing radical mastectomy, those undergoing simple mastectomy, and those undergoing a lumpectomy followed by radiation therapy (17,18).

Use of Factorial Designs

In view of the cost and feasibility issues of large, simple clinical trials, one technique to improve efficiency is to test two or more hypotheses simultaneously in a single trial, using a factorial design. In a 2x2 factorial design, participants are first randomized to treatments A or B to address one hypothesis, and then within each treatment group there is further randomization to treatments C or D to evaluate a second question. In a 2x2x2 factorial design, each of these subgroups would be further randomized into two additional intervention groups to address a third hypothesis, and so on.

For example, ISIS-2 used a 2x2 factorial design to evaluate streptokinase as well as aspirin in the

treatment of acute MI (33). Patients were randomized to 1.5 million units of streptokinase or a placebo given intravenously, as well as 160 mg of aspirin or a placebo daily for 30 days, for a total of four treatment groups: participants receiving either streptokinase alone, aspirin alone, both active agents, or both placebos.

The principal advantage of the factorial design is that it allows the simultaneous testing of more than one question in a single trial, while costing little more than a trial of one of the questions alone. Ideally, of course, the additional treatments in a factorial design should not complicate trial operations, materially affect eligibility requirements, or cause side effects that could lead to poor compliance or losses to followup. In addition, the possibility of an interaction between the treatment regimens must be considered. Although the possibility of such interactions is considered by some to be a limitation of a factorial design, only through such a design can any combined effects of trial treatments be detected (43). ISIS-2 showed that streptokinase alone and aspirin alone clearly reduced 35-day vascular mortality, but that the participants who received both drugs experienced the greatest reduction in risk. The factorial design allowed this interaction to be assessed, which would not have been possible in a single-factor study.

Subgroup Analyses from Randomized Trials

Looking at the effect of an intervention among specific subgroups of participants might appear to be a way to address the question of whether the findings of a trial conducted in a wide group of patients are also applicable to patients with particular characteristics. It is important to keep in mind, however, that the most valid comparison in a randomized trial is between the originally allocated treatment groups. It is only in this comparison that randomization, in trials of adequate sample size, assures nearly even distribution of all the potential confounding variables, both known and unknown. In an analysis of any subgroup, whether defined on the basis of compliance or any other baseline characteristic, the comparison is no longer randomized and the potential role of confounding must be evaluated and controlled to the extent possible, just as in any observational study.

This point is illustrated by the experience of the Coronary Drug Project trial, a study testing the effect of the cholesterol-lowering drug clofibrate in the reduction of mortality following MI (10). In that trial, the five-year mortality rates in the groups receiving the clofibrate and the placebo were very similar (18.0 percent versus 19.5 percent). Because there was substantial noncompliance with the clofibrate regimen, the investigators attempted to more clearly evaluate the efficacy of the drug by also analyzing the mortality rates within the clofibrate group. They found that patients whose compliance was at least 80 percent had a mortality rate of 15 percent, compared with a rate of 24.6 percent among those who were less compliant. Such a finding might be erroneously interpreted to indicate that clofibrate reduces mortality. An analysis within the placebo group, however, found a similar disparity in mortality between compliers and noncompliers, with rates of 15.1 percent and 28.2 percent, respectively. Even after controlling for 40 known possible confounding variables, researchers still found a difference between the mortality rates of compliers and noncompliers in the placebo group. These data indicate that subgroup analyses of compliers did not provide valid results, because of the inability to control for the confounding effects of differences between compliers and noncompliers that independently affected their prognosis.

Another problem in the interpretation of findings from subgroup analyses relates to the meaning of testing for statistical significance, or the p-value. In medical research, the conventional level of statistical significance is $p=0.05$. This means that once in 20 times, a finding will be said to be statistically significant by chance alone, even though no difference between the treatment groups actually exists. This implies that if enough comparisons were made in a trial, as would occur in an evaluation of the treatment's effect among a large number of subgroups, one in 20 would be statistically significant, even if the intervention

actually had no effect. The interpretation of this finding, however, greatly depends on whether the subgroup was previously defined as being of interest or was found by "fishing" or "data-dredging."

To illustrate the potential pitfalls of analyzing many subgroups, the ISIS-2 investigators carried out analyses according to patients' astrological signs. The researchers found that those born under the birth signs Gemini and Libra experienced a nonsignificant adverse effect of aspirin in acute MI, whereas aspirin significantly reduced the mortality rate of those born under all other zodiac signs (33). This was not an a priori hypothesis, as there was no clinical basis for the belief that aspirin would differentially affect those born under certain astrological signs, and the analysis clearly demonstrates the caution that must be used in interpreting the results of subgroup analyses.

The Decision To Terminate a Trial Early

In the design phase of a trial, the researchers need to develop guidelines for deciding whether the trial should be modified or terminated before its originally scheduled conclusion. To assure that the welfare of the participants is protected, the unblinded data should be monitored by a group that is independent of the investigators who are conducting the trial. If the data indicate that the intervention has a clear and extreme benefit on the primary endpoint, or if a treatment is clearly harmful, the modification or early termination of the trial must be considered.

A decision to terminate a study early is based on a number of complex issues and must be made with a great deal of caution. It is critical that a trial not be stopped prematurely based solely on emerging trends from a small number of patients, because these findings might well be transient and disappear or even reverse after data have accumulated from a larger sample. As a general rule, the first requirement for even considering the modification or early termination of an ongoing trial is the observation of a sustained statistical association that is so extreme, and so highly statistically significant, that its emergence by chance alone would be virtually impossible. The observed association must then be considered in the context of the totality of evidence. A number of specific guidelines have been used in various studies, but the aim is to achieve an equitable balance between, on the one hand, protecting randomized participants from real harm and, on the other hand, minimizing the risks of mistakenly modifying or stopping the trial prematurely.

Whenever a trial is ended prematurely because of findings related to one endpoint, the ability to answer other, often equally important, questions may be lost. The Physicians' Health Study is a case in point. The study was designed to evaluate two primary prevention hypotheses: whether low-dose aspirin reduces cardiovascular mortality and whether beta-carotene decreases cancer incidence. In early 1988, the trial's Data Monitoring Board prematurely terminated the randomized aspirin component of the trial (7a). This decision was based on all the available evidence, including three major considerations: the presence of a statistically significant ($p<0.00001$) reduction in the risk of MI among those in the group receiving aspirin; the fact that no effect of aspirin on cardiovascular mortality could be detected in the trial until the year 2000 or later, because of the exceptionally low cardiovascular death rates among the participating physicians; and the fact that aspirin was subsequently prescribed for more than 85 percent of the participants who experienced nonfatal MIs, which would render any later findings about aspirin and cardiovascular mortality particularly difficult to interpret.

Two other significant outcomes of interest in relation to aspirin were stroke and cardiovascular death, both of which occurred less frequently than MI. As a result of the early termination of the aspirin component, participants experienced inadequate numbers of strokes and cardiovascular deaths to permit a reliable assessment of aspirin's effect. There was an apparent increased risk of stroke—primarily in the subgroup of hemorrhagic strokes—but it was not statistically significant. No reduction in the risk of cardiovascular mortality was associated with aspirin. Although a num-

ber of explanations have been proposed, the primary consideration must be that the number of cardiovascular deaths in the trial at the time the aspirin component was terminated was simply too small to reliably evaluate the endpoint. Thus, two major pieces of the benefit-to-risk equation for the use of aspirin in the primary prevention of cardiovascular disease could not be determined, because of the ethical and practical considerations that prompted the early termination of the aspirin component of the trial.

Role of Meta-Analyses in Randomized Trials

The sample size of a trial and its resultant statistical power determine the extent to which chance may have influenced the study findings. If a study's sample size is inadequate, then a finding of no statistically significant association between the intervention and the outcome (a so-called null finding) may well be uninformative, because a true lack of association would be difficult or impossible to distinguish from a true association that simply could not be detected because of inadequate statistical power.

The ambiguity of the results from individual trials with small samples provides a strong rationale for much larger trials that could reliably detect modest treatment effects. Some investigators have argued that small trials should be pursued first, with larger investigations undertaken only if shown to be necessary. Because uninformative null results from small trials may erroneously suggest no effect, however, it would appear far preferable to mount a large trial once there is sufficient belief in a treatment's potential. If the effect is far greater than anticipated, a large trial can always be terminated earlier than scheduled.

Although a single well-designed and -conducted trial of sufficient size to detect the true effects of an intervention is usually considered optimal, in the absence of a definitive study, statistical overviews or meta-analyses that consider in aggregate the data from several small trials can provide useful information by minimizing the role of chance as an explanation for the findings (30) (see M.P. Longnecker, *Meta-Analysis,* background paper no. 4). However, meta-analysis cannot overcome the effects of bias or confounding present in the individual trial results.

One of the most important uses of meta-analyses of small trials may be not to provide a definitive answer to a question, but to provide a reliable estimate of the most likely effect of an intervention. That estimate then can be used in planning a future trial with adequate power to detect such an effect if it truly exists. With respect to estimating the size of any reduction in risk, the results from a pilot study or even a single small trial are likely to be quite unstable due to sampling variability. By contrast, the unique strength of meta-analysis of data from all randomized trials is to minimize the variability of the overall estimate that is obtained from each individual study. Thus, meta-analyses provide the most reliable risk estimates that can be obtained in the absence of individual trials of adequate statistical power.

The pitfalls of estimating the likely effects of an intervention from a single small trial instead of an overview can be illustrated by comparing two trials that tested the effects of beta-blockers, drugs given in the early acute phase of a heart attack: the Metoprolol in Acute Myocardial Infarction (MIAMI) trial and the First International Study of Infarct Survival (ISIS- 1). Before the initiation of the MIAMI trial, the investigators conducted a pilot study of approximately 1,400 participants. Based on an observed 36-percent reduction in total mortality, approximately 6,000 individuals were enrolled in the full-scale trial. By contrast, the sample size for ISIS- 1 was calculated from an overview of 21 previously conducted trials of beta-blocker therapy, which indicated an approximate 10-percent reduction in total mortality. On the basis of this estimate, more than 16,000 patients were enrolled in ISIS-1. When the two trials were completed, the estimates of the effects of treatment were similar, with vascular mortality reduced by approximately 13 percent in MIAMI and 15 percent in ISIS- 1, but the results from the MIAMI trial did not achieve statistical significance, while the results from ISIS-1 did (32,39).

CONCLUSIONS

For addressing many important research questions, randomized trials are neither necessary nor desirable. However, if the effects of a hypothesized intervention are likely to be only small to moderate in size, a trial with a large sample will be necessary to provide a definitive test of such research questions. Large trials can, if properly designed, be conducted using relatively simple protocols in which minimal screening or data collection is required.

There must be flexibility in the application of these principles to suit the particular circumstances of each research question. Although trials of in-hospital treatment of acute MI may be extremely simple in design and collect minimal data, those of chronic disease treatments, such as the digitalis trial in congestive heart failure, may require somewhat more involved protocols. Trials of primary prevention, in turn, may necessitate more prolonged screening phases to enroll populations of willing and eligible participants, and longer treatment and followup will be needed to accrue sufficient numbers of endpoints to permit valid tests of the trial hypotheses.

All these trials are characterized by their emphasis on enrolling large numbers of participants; testing an intervention's effect on a readily ascertained, clinically important outcome; and collecting a relatively limited amount of baseline and followup data. Many areas of medical research are suitable for large, simple trial protocols. Some types of questions, however, do not lend themselves to testing using such an approach, and are therefore best evaluated in more traditional trials with complex protocols and extensive data collection.

Nevertheless, many questions could be tested using far simpler protocols than those that have been used in most randomized controlled trials. Physicians are by training-and perhaps by temperament-oriented toward gathering detailed information on individual patients. This has very likely contributed to the use of very complex trial protocols with small samples in which hundreds of variables are collected on a few participants. Although an extensive history on a particular patient is correctly viewed as crucial to rendering individually appropriate clinical decisions, the same is not necessarily true for attempting to answer fundamental questions regarding the effectiveness of a promising medical intervention. It is far preferable, in the case of many trials, to have data on a few variables for hundreds or thousands of participants than to collect information on scores of variables from only a few study participants.

Where appropriate, therefore, large and simple trials can provide more reliable tests of an intervention than can other feasible research approaches. As their broad contributions to medicine become more widely understood, such investigations may play an increasing role in answering important research questions, and in providing a sound basis for formulating rational clinical decisions for individual patients and public health policy for the general population.

REFERENCES

1. Aboulker, J.P., and Swart, A. M., "Preliminary Analysis of the Concorde Trial," *Lancet* 341:889-890, 1993.
2. Altman, L., "Fall of a Man Pivotal in Breast Cancer Research," *The New York Times,* p. B1O, Apr. 4, 1994.
3. Boissel, J. P., "Impact of Randomized Clinical Trials on Medical Practices," *Controlled Clinical Trials* 10:12S-134S, *1989.*
4. Buring, J.E., and Hennekens, C. H., "Cost and Efficiency in Clinical Trials: The U.S. Physicians' Health Study," *Statistics in Medicine* 9:29-33, *1990.*
5. Buring, J. and Hennekens, C. (for the Women's Health Study Research Group), "The Women's Health Study: Summary of the Study Design," *Journal of Myocardial Ishcemia* 4:27-29, 1992.
6. Byar, D., "Design Considerations for AIDS Trials," *Journal of AIDS* 3(suppl. *2):* S16-*S19,* 1990.
7. Byar, D. P., Schoefeld, D. A., Green, S. B., et al., "Design Considerations for AIDS

Trials," *New England of Medicine 323: 1343-1348, 1990.*

7a. Cairns, J. Cohen, L., Colton, T., et al., "Issues in the Early Termination of the Aspirin Component of the Physicians' Health Study, *Annals of Epidemiology* 1:345-405, *1991.*

8. CLASP (Collaborative Low-Dose Aspirin Study in Pregnancy) Collaborative Group, "CLASP: A Randomised Trial of Low-Dose Aspiring for the Prevention and Treatment of Pre-Eclampsia Among 9,364 Pregnant Women," *Lancet* 343:619-629, *1994.*

9. Concorde Coordinating Committee, "Concorde: MRC/ANRS Randomised Double-Blind Controlled Trial of Immediate and Deferred Zidovudine in Symptom-Free HIV Infection, " *Lancet* 343:871-881, 1994.

10. Coronary Drug Project Research Group, "Influence of Adherence to Treatment and Response of Cholesterol on Mortality in the Coronary Drug Project, New *England Journal of Medicine 303: 1038, 1980.*

11. Crile, G., "Treatment of Breast Cancer by Local Excision," *American Journal of Surgery* 109:400-403, *1965.*

12. Digitalis Investigation Group, Trial To Evaluate the Effect of Digitalis on Mortality in Heart Failure, final protocol, U.S. Department of Health and Human Services, National Institutes of Health, National Heart, Lung and Blood Institute, Bethesda, MD, unpublished report, 1992.

13. Ellenberg, S. S., "Do Large, Simple Trials Have a Place in the Evaluation of AIDS Therapies? *Oncology* 6:55-59, 1992.

14. Ellenberg, S. S., and Foulkes, M. A., "The Utility of Large, Simple Trials in the Evaluation of AIDS Treatment Strategies, *Statistics in Medicine* 13:405-413, 1994.

15. Ellenberg, S. S., Cooper, E., Eigo, J., et al., "Studying Treatments for AIDS: New Challenges for Clinical Trials-A Panel Discussion at the 1990 Annual Meeting of the Society for Clinical Trials," *Controlled Clinical Trials* 13:272-292,1992.

16. EMMES Corporation, "Reanalysis of NSABP Protocol B-06—Final Report," prepared by D.M. Stablein and staff of the EMMES Corporation, Potomac, MD, Apr. 11, 1994.

17. Fisher, B., Bauer, M., Margolese, R., et al., "Five-Year Results of a Randomized Clinical Trial Comparing Total Mastectomy and Segmental Mastectomy with or without Radiation in the Treatment of Breast Cancer," New *England Journal of Medicine* 312:665-673, *1985.*

18. Fisher, B., Redmond, C., Fisher, E.R., et al., "Ten-Year Results of a Randomized Clinical Trial Comparing Radical Mastectomy and Total Mastectomy with or without Radiation," New *England Journal of Medicine* 312:674-681, *1985.*

19. Fishl, M.A., Richman, D. D., Grieco, M. H., et al., and the AZT Collaborative Working Group, "The Efficacy of Azidothymidine (AZT) in the Treatment of Patients with AIDS and AIDS-Related Complex: A Double-Blind, Placebo-Controlled Trial," New *England Journal of Medicine 317: 185-191, 1987.*

20. Francis, T., Kerns, F.T., Voight, R. B., et al., "An Evaluation of the 1954 Poliomyelitis Vaccine Trials: Summary *Report, ''American Journal of Public Health* 45:1-49, *1955.*

21. *Garg,* R., Scientific Project Officer, U.S. Department of Health and Human Services, National Institutes of Health, National Heart, Lung and Blood Institute, Bethesda, MD, personal communication, Jan. 6, 1992.

22. Gibbons, R.J., Holmes, D.R., Reeder, G. S., et al., "Immediate Angioplasty Compared with the Administration of a Thrombolytic Agent Followed by Conservative Treatment for Myocardial Infarction," New *England Journal of Medicine* 328:684-691, *1993.*

23. Grines, C.L., Browne, K.F., Marco, J., et al., "A Comparison of Immediate Angioplasty with Thrombolytic Therapy for Acute Myocardial Infarction," New *England Journal of Medicine* 328:673-679, *1993.*

24. Gruppo Italiano por 10 Studio della Streptochinasi nell'Infarto Miocardico (GISSI),

"Effectiveness of Intravenous Thrombolytic Treatment in Acute Myocardial Infarction," *Lancet* i:397-402, 1986.

25. Gruppo Italiano por 10 Studio della Sopravvivenza nell'Infarto Miocardico, "GISSI-2: A Factorial Randomized Trial of Alteplase versus Streptokinase and Heparin versus no Heparin among 12,490 Patients with Acute Myocardial Infarction," *Lancet* 336:65-71, 1990.
26. GUSTO Investigators, "An International Randomized Trial Comparing Four Thrombolytic Strategies for Acute Myocardial Infarction," New *England Journal of Medicine* 329:673-682, *1993.*
27. Hennekens, C. H., and Buring, J. E., *Epidemiology in Medicine* (Boston, MA: Little, Brown, and Company, 1987).
28. Hennekens, C. H., and Buring, J. E., "Methodologic Considerations in the Design and Conduct of Randomized Trials: The U.S. Physicians' Health Study," *Controlled Clinical Trials* 10(suppl.): 142S-150S, 1989.
29. Hennekens, C. H., and Eberlein, K., "A Randomized Trial of Aspirin and Beta-Carotene Among U.S. Physicians, *Preventive Medicine 14: 165-168, 1985.*
30. Hennekens, Buring, J. E., and Hebert, P., "Implications of Overviews for Randomized Trials," *Statistics in Medicine* 6:397-402, *1987.*
31. Hennekens, C. H., Jonas, M. A., and Buring, J. E., "The Benefits of Aspirin in Acute Myocardial Infarction: Still a Well-Kept Secret in the U.S.," *Archives of Internal Medicine* 154:37-39, *1994.*
32. ISIS-1 Collaborative Group, "ISIS-l: A Randomised Trial of Intravenous Atenolol Among 16,027 Cases of Suspected Acute Myocardial Infarction," *Lancet* 2:57-66, 1986.
33. ISIS-2 Collaborative Group, "ISIS-2: Randomised Trial of Intravenous Streptokinase, Oral Aspirin, Both or Neither Among 17,187 Cases of Suspected Acute Myocardial Infarction," *Lancet* 2:349-360, *1988.*
34. *ISIS-3* Collaborative Group, "ISIS-3: A Randomised Trial Comparing SK vs. tPA vs. APSAC and Comparing Aspiring Plus Heparin vs. Aspirin Alone Among 41,299 Suspected Acute Myocardial Infarction," *Lancet* 339:1-18, 1992.
35. Lamas, G. A., Pfeffer, M.A., Hamrn, P., et al. (for the SAVE Investigators), "Do the Results of Randomized Clinical Trials of Cardiovascular Drugs Influence Medical Practice?" New *England Journal of Medicine* 327:241-247, *1992.*
36. Lang, J. M., Buring, J. E., Rosner, B., et al., "Estimating the Effect of the Run-In on the Power of the Physicians' Health Study," *Statistics in Medicine* 10: 1585-1593, 1991.
37. Lange, R. A., and Hillis, L. D., "Immediate Angioplasty for Acute Myocardial Infarction," New *England Journal of Medicine* 328:726-728, *1993.*
38. Meier, P., "Polio Trial: An Early Efficient Clinical Trial," *Statistics in Medicine* 9:13-16, *1990.*
39. The MIAMI Trial Research Group, "Metoprolol in Acute Myocardial Infarction (MIAMI): A Randomised Placebo-Controlled International Trial," *European Heart Journal 6: 199-226, 1985.*
40. Pfeffer, M.A., Moye, L. A., Braunwald, E., et al., "Selection Bias in the Use of Thrombolytic Therapy in Acute Myocardial Infarction," *Journal of the American Medical Association* 266:528-532, *1991.*
41. Ridker, P. M., O'Donnell, C., Marder, V., et al., "Large-Scale Trials of Thrombolytic Therapy for Acute Myocardial Infarction: GISSI-2, ISIS-3, GUSTO- l," *Annals of Internal Medicine* 119:530-532, 1993.
42. Smith, R. P., Meier, P., Abrams, D., et al., "Time for Some Really Large, Simple Randomized Trials in HIV Disease," *Oncology* 6:63-64, 1992.

43. Stampfer, M., Buring, J. E., Willet, W., et al., "The 2x2 Factorial Design: Its Application to a Randomized Trial of Aspirin and Carotene in U.S. Physicians," *Statistics in Medicine 4:11* 1-116, 1985.

44. The Steering Committee of the Physicians' Health Study Research Group, "Final Report on the Aspiring Component of the Ongoing Physicians' Health Study," New *England Journal of Medicine* 321: *129-135,* 1989.

45. Tognoni, G., Franzonsi, M. G., Garattini, S., et al., "The Case of GISSI in Changing the Attitudes and Practices of Italian Cardiologists," *Statistics in Medicine 9: 17-27, 1990.*

46. Topol, E.J., "Thrombolytic Intervention," *Textbook of Interventional Cardiology,* E.J. Topol (cd.) (Philadelphia, PA: W.B. Saunders, 1989).

47. Volberding, P. A., Lagakos, S.W., Koch, M. A., et al., and the AIDS Clinical Trials Group of the National Institute of Allergy and Infectious Diseases, "Zidovudine in Asymptomatic Human Immunodeficiency Virus Infection: A Controlled Trial in Persons with Fewer Than 500 CD4-Positive Cells Per Cubic Millimeter," New *England Journal of Medicine* 322:941-949, *1990.*

48. Wilhelmsen, L., "The Applicability of the Results of Streamlined Trials to Clinical Practice," *Statistics in Medicine 9: 185-191, 1990.*

49. Yusuf, S., Collins, R., and Pete, R., ''Why Do We Need Some Large, Simple Randomized Trials?" *Statistics in Medicine* 3:409-420, *1984.*

50. Yusuf, S., Collins, R., Pete, R., et al., "Intravenous and Intracoronary Fibrinolytic Therapy in Acute Myocardial Infarction: Overview of Results on Mortality, Reinfarction, and Side Effects from 33 Randomised Controlled Trials," *European Heart Journal* 6:556-585, *1985.*

51. Yusuf, S., Garg, R., Held, P., et al., "Need for a Large Randomized Trial To Evaluate the Effects of Digitalis on Morbidity and Mortality in Congestive Heart Failure," *American Journal of Cardiology* 69:64G-70G, *1992.*

52. Yusuf, S., Held, P., Tee, K. K., et al., "Selection of Patients for Randomized Controlled Trials: Implications of Wide or Narrow Eligibility Criteria," *Statistics in Medicine* 9:73-86, *1990.*

53. Zijlstra, F., de Boer, M. J., Hoomtje, J.C.A., et al., "A Comparison of Immediate Coronary Angioplasty with Intravenous Streptokinase in Acute Myocardial Infarction, ''New *England Journal of Medicine* 328:680-684, *1993.*