Statistical Science
1991, Vol. 6, No. 1, 52–88

Ethics and Statistics in Randomized Clinical Trials

Richard M. Royall

Abstract. Randomized clinical trials can present a scientific/ethical dilemma for clinical investigators. Statisticians have tended to focus on only one side of this dilemma, emphasizing the statistical and scientific advantages of randomized trials. Here we look at the other side, examining the personal care principle on which the physician-patient relationship is based and observing how that principle can make it difficult or impossible for a physician to participate in a randomized clinical study. We urge that the view that randomized clinical trials are the only scientifically valid means of resolving controversies about therapies is mistaken, and we suggest that a faulty statistical principle is partly to blame for this misconception. We conclude that statisticians should be more sensitive to the physician's responsibility to the individual patient and should, besides promoting randomized trials when they are ethically and practically feasible, work to improve the planning, execution, and analysis of nonrandomized clinical studies.

Key words and phrases: Nonrandomized control group, observational study, physician-patient relationship, randomization, selection bias.

1. INTRODUCTION

> . . . from my associations with doctors in controlled trials I have learned that the better the statistician understands the doctor/patient relationship and the doctor's very real and unique ethical problem the better he can help to devise a trial that may be less than ideal experimentally but yet likely to be of some, and perhaps considerable, value to medicine. . . . still more important, I have learned that though the statistician himself may never see a patient . . . nevertheless, he cannot sit in an armchair, remote and Olympian, comfortably divesting himself of all ethical responsibility. As a partner in a combined endeavor a full share of that responsibility will always lie with him . . . (Hill, 1963).

Randomized clinical trials are widely regarded as the most scientifically sound approach to determining which of two medical treatments is better. Some argue that they are *essential* for the advancement of medicine. For example, Zelen (1979), states that disruption of such trials would undermine ". . . the scientific basis of therapeutic medicine," Tukey (1977) expresses the conviction that ". . . the only source of reliable evidence about the usefulness of almost any sort of therapy or surgical intervention is that obtained from well-planned and carefully conducted randomized . . . clinical trials," and Cowan (1981) writes "With some exceptions, participation of any group of patients in a nonrandomized trial is wholly unjustified and unethical since nothing can be learned from it." While not all statisticians would endorse such strong judgments, most would agree with Byar et al. (1976) that "Randomized clinical trials remain the most reliable method for evaluating the efficacy of therapies."

Despite the enthusiasm that they generate on scientific and statistical grounds, many randomized clinical trials are beset by reservations and discomfort arising from ethical considerations. Here we will examine the ethical and statistical arguments bearing on the dilemma faced by the clinical investigator. A trial that compared two therapies for newborn infants in severe respiratory distress will be used as an example. This example is chosen, not as a particularly extreme instance of the general

Richard M. Royall is Professor, Department of Biostatistics, The Johns Hopkins University, 614 N. Wolfe Street, Baltimore, Maryland 21205.

problems discussed here, but because its authors were unusually candid in describing their beliefs going into the study and their rationale for proceeding as they did.

2. THE PHYSICIAN'S OBLIGATION TO THE PATIENT

2.1 The Personal Care Principal

"The traditional concept of the physicians' relation to his patient is one of unqualified fidelity to that patient's health" (Fried, 1974, page 50). Schafer (1982) notes that, in his traditional role of healer, the physician's commitment is "exclusively and unequivocally to promote the interests of his patient." This concept has no accepted name; it is very closely related to what Veatch (1981) calls the "Hippocratic principle" and Marquis (1983) names the "therapeutic obligation." I will follow Fried in calling it the personal care principle. It relates specifically to the professional relationship between physician and patient and expresses (Chapman, 1984) the physician's obligation "to place the patient's interests before all else within the professional relationship."

The personal care principle is commonly stated as the physician's duty "to do what is best for the patient." It is not entirely clear whether it is implied to the Hippocratic Oath ("Whatever houses I may visit, I will come for the benefit of the sick . . . "), but the Physician's Oath adopted by the World Medical Association at Geneva (1948) is explicit: "the health of my patient will be my first consideration." In the context of clinical trials, the principle appears in the Medical Research Council of Great Britain's (1964) statement on Responsibility in Investigations on Human Subjects: "It goes without question that any doctor taking part in . . . a collective controlled trial is under an obligation to withdraw a patient from the trial, and to institute any treatment he considers necessary, should this, in his personal opinion, be in the better interests of his patient." And individual authors writing about clinical trials routinely acknowledge " . . . the fundamental principle that the physician-investigator's primary responsibility is to his patient" (Shaw and Chalmers, 1970).

In addition to whatever moral force might support the personal care principle, there is a practical need for it—its absence would create severe obstacles to the proper functioning of the present system of medical care. One who doubts that his physician's professional advice is motivated by what is best for the patient is effectively cut off from adequate care. Blind faith that no physician ever put another interest above the health of a patient is not necessary or even desirable. But confidence that the personal care principle is usually honored seems essential if people are to be motivated to seek out needed medical help and to follow prescribed therapies.

2.2 Paternalism and Autonomy

Views about the proper relationship between physician and patient are not static. Before the 1940s the dominant image was a paternalistic one in which the focus was on the physician's responsibility to do what, in his judgment, was best for the patient. Recent decades have seen increasing emphasis on the rights of the autonomous patient, particularly the rights to be informed and to make his or her own decisions. The personal care principle does not demand that the physician does whatever he thinks is best for the patient, regardless of the patient's preferences: the role of the physician is to advise and propose, not to impose his judgments. The right to decide what is truly in the patient's best interests belongs to the patient. As Chapman (1984) put it "There are times when the doctor does indeed know best, even though he or she may not impose a decision; ultimate choice is uncontestably the patient's." This view is exemplified in recent legal judgments that the decision of a patient whose religious beliefs forbid blood transfusion must be respected, even by a physician who is convinced of the urgent medical need for the transfusion. Of course many patients choose to play a passive role, trusting the physician to do "Whatever you think is best." And of course the basis for this decision is the assumption that the personal care principle will be respected.

2.3 Competence

The personal care principle is not the physician's only rule of professional ethics. Another obligation relates directly to our discussion: the physician must maintain competence. His actions are ethically sound only if they are directed by both the personal care principle and adequate professional knowledge and skills. A physician does not face an obligation to provide the best available therapy for his patient. His obligation is to provide what is best in his judgment, and to ensure that his judgment is medically competent. No more is humanly possible. If two competent clinicians, both observing the personal care principle, disagree about the relative merits of two therapies, then one may be using what is in fact an inferior treatment, but neither is behaving unethically.

2.4 Limitations to Personal Care

There are certain situations in which the physician's responsibility to serve the interests of the patient is restricted on behalf of the interests of society. For example, although the physician is generally charged with maintaining confidentiality, he is required to report gunshot wounds to the police. Clearly such a report might be contrary to the interests of a wounded fugitive. For another example, Food and Drug Administration regulations limit the conditions under which many drugs can be used, so that a physician may not be allowed to prescribe what he or she thinks would be best for the individual patient. It is important to note that these situations, in which the personal care principle is compromised for a presumed social good, are explicitly circumscribed by public law and not left to the personal judgment of the physician.

3. RANDOMIZED CLINICAL TRIALS

Clinical trials are medical experiments with patients as subjects. We will consider trials that take place when there is uncertainty about which of two therapies is preferable. Some patients receive one treatment and some the other, then results in the two groups are compared. The defining characteristic of *randomized clinical trials* (RCTs) is that each patient's therapy is determined by a specified chance mechanism: each patient is *randomized* to treatment A or treatment B. Questions of ethics arise because the treatment that an individual patient receives is determined, not by the physician acting under the personal care principle, but by the randomizing device. Before looking at these ethical questions, let us consider why we might want to carry out an RCT.

3.1 Advantages of RCTs

We want to learn about these two treatments, A and B. If we see different results in the A group than in the B group, we want to attribute this to differing treatment effects. But of course this conclusion can be invalidated by the observation that the two groups of patients were not comparable before treatment. Randomization precludes one notorious source of noncomparability, selection bias. It ensures that the two groups cannot differ because the physician tended (either consciously or not) to select treatment A for the more severely ill patients, for example.

Besides protecting against selection bias, randomization also tends to balance the treatment groups with respect to covariates. For example, it tends to produce groups that are similar with respect to the distribution of prognostic factors. Of course randomization does not *guarantee* to produce groups that are similar, say in their age distribution, and if it is found that by chance the groups are different, then statistical adjustments can be made. But one of the most important advantages of randomization relates to unobserved, perhaps unknown, covariates, for which statistical adjustments are impossible. Although it cannot be known for certain that the groups in the study are in fact comparable on such an unobserved covariate, X, the tendency of randomization to produce balanced groups can ensure that severe imbalance on X is improbable, thus strengthening the argument that an observed difference between treatment groups is due to the treatments, not to X-differences between the groups.

The third advantage of randomization that is commonly cited (e.g., Byar et al., 1976) relates to the probabilities used in statistical analysis. Probabilities deliberately introduced via random treatment assignment can be used in analyzing the results, and such analyses are then surely valid in a sense that cannot be guaranteed for analyses based on probabilistic assumptions not grounded in the act of randomization.

Of course RCTs have some disadvantages, too. For example, obtaining the patients' informed consent to participate in a study can be made more difficult by the need to explain randomization. Besides prolonging the investigation, this can produce a study population that differs markedly in education and attitudes from the target population. However, the advantages of RCTs are such that few would disagree with the position that, other things being equal, a comparative trial is strengthened by the judicious use of randomization.

3.2 Two Types of RCTs

An RCT that is carried out by clinicians who are themselves quite unsure which treatment is better will be called an *experimental* trial. Many trials are of a different sort: their authors are confident that one treatment is superior. Their purpose is to demonstrate that superiority in a way that will be convincing to others. These will be called *demonstration* trials. Such trials are often undertaken by the developers of a new therapy who, having used it successfully in a series of patients, believe that it is better than the old treatment. But sometimes, as in proposed trials of certain controversial cancer therapies, the purpose is to provide a convincing demonstration of the inferiority of an unconventional treatment (Mulder, 1981; Cowan, 1981).

Demonstration trials clearly present ethical prob-

lems, because they involve a physician's assigning some of his patients to a treatment that he believes to be inferior. What can be said in their defense? There are two arguments. One of these looks at the beneficial consequences of an RCT and weighs the cost against these benefits; the other is more philosophical, centered on the definition of "knowledge." We describe the first argument as practical and the second as epistemological.

3.3 The Practical Argument for Demonstration Trials

The practical argument is based on the utilitarian objective of maximizing the total good. It acknowledges that there is a cost, that some patients in the demonstration trial will not receive what, in their physician's best judgment, is best for them. But it argues that this cost is exceeded by the benefits to many more patients whose physicians will have been convinced by the demonstration trial not to use the inferior treatment.

This utilitarian argument seems to be particularly attractive to statisticians and others who are not bound by the clinician's obligation to the individual patient. However, according to the personal care principle, the patient has a right to expect that his physician will act in his best interests and will not sacrifice them on behalf of other patients. Although the physician might judge that one patient's interests are less valuable than those of others, the principle forbids his making the trade. Specifically, it forbids his giving a patient what he believes to be an inferior treatment in order to demonstrate its inferiority for the benefit of others.

3.4 The Epistemological Argument for Demonstration Trials

The practical argument for demonstration trials emphasizes their benefits, attempting to show that they are great enough to outweigh the costs. On the other hand, the epistemological argument claims that there is actually no cost. According to this argument, if therapy A were known to be better than B, then there would be no need for an RCT. And if it is not known, then the physician who believes that A is better has no sound basis for recommending A; his belief represents only a personal opinion, an unscientific hunch, and is not a proper basis for responsible professional judgment. Thus even though in randomly assigning half of his patients to therapy B he fully believes that he is doing less than he can for those patients, there is actually no sacrifice made, no ethical cost incurred.

Freund (1982) used this argument in urging that more clinical trials should be done in surgery.

> While the ... new technique may seem quite different or novel to the general medical and surgical community, in the eyes of those involved in its evolution there is no such feeling. A pilot study of several patients will have been performed with success.. The surgeons primarily involved... may be enthusiastic in their advocacy of the new procedure. To submit their patients to a large clinical randomized study is not ethical in their opinion, since they have already seen that their new approach is better... and yet, in terms of statistically significant, unbiased, truly tested knowledge, a sound scientific database does not exist.
>
> ...much of what the surgeon assumes he knows is not based on solid scientific data, but rather on training, experience, and reinforcement. The choice of treatment is neither more nor less likely to be correct if made arbitrarily than if assigned randomly in the clinical trial. The two courses of action can thus be considered ethically equivalent in terms of patient risk.
>
> ...even an opinion held with strong conviction is not a sufficient basis for ethical action; passionate opinion does not make an incorrect opinion into a correct one.

This echoes the argument of Byar et al. (1976) who proposed that when the relative merits of two therapies are in dispute an RCT "...involves no ethical compromise concerning the patients who will be involved in the trial in that no patient will receive a therapy that is known to be inferior to another."

Freedman (1987) presents a variation on this argument. He considers the situation where "There exists (or, in the case of novel therapy, there may soon exist) an honest, professional disagreement among expert clinicians about the preferred treatment." Now some investigators who "have a decided preference for [therapy] B but wish to conduct a trial" appear before an ethics committee:

> The ethics committee asks the investigators whether, if they or members of their families were within [the study] population P, they would not want to be treated with their preference B? An affirmative answer is often thought to be fatal to the prospects for such a trial, yet the investigators answer in the affirmative. Would [such] a trial... be ethical?
>
> I believe that it clearly is ethical... The ethics of medical practice grants no ethical or normative meaning to a treatment preference, however powerful, that is based on a hunch or

on anything less than evidence publicly presented and convincing to the clinical community.

The epistemological argument undermines the personal care principle. It does this not by direct contradiction, but by emptying the class of situations to which the principle applies. According to that argument, it might be true that whenever the physician *knows* therapy A to be better than B, he must try to provide A. But when he doesn't *know*, when he only *believes* A to be better, he is under no such obligation.

So when does the personal care principle apply? When does he *know*? Is it when his beliefs agree with some broad professional consensus? There are too many examples where the consensus has proved wrong to accept this definition. The consensus is certainly relevant in making judgments about competence, and one whose beliefs are at odds with the majority view must be prepared to give rational arguments to support his position. But knowledge is not simply whatever is widely believed by experts. A dissenter can be the one who "knows," while the "well-known fact" can prove to be only a popular misconception.

If the consensus of experts does not define the physician's "knowledge," then what does? How can we decide when he "knows" what is best, and is therefore bound by the personal care principle? Does the physician "know" that A is the better therapy when an RCT has reached this conclusion? Of course not. Structuring a study as an RCT does not ensure that it will not reach an erroneous conclusion; it only provides a way to measure and control the *probability* of such an outcome. An RCT can provide valuable evidence, which can justify strong beliefs. It can make one very confident and convince others as well. It cannot prove anything.

The physician never knows; he has only his informed judgment, his beliefs, and it is on these that he must act in his patient's interest. If the personal care principle applied only when he *knew* what was best, it would be meaningless.

3.5 Unacceptability of Demonstration Trials

The personal care principle prohibits demonstration trials. The patient has a right to expect that the physician will not deliberately provide what he believes to be inferior care, even for the worthy purpose of generating sound scientific evidence that his belief is correct.

How then is the advocate of a new or improved therapy to demonstrate the procedure's superiority in a way that other clinicians will find convincing? I will argue later that alternatives to RCTs should be more highly valued. But RCTs provide the most convincing evidence, and if demonstration trials are forbidden because they entail violation of patients' individual rights, then science will suffer. And, to the extent that science is used for the betterment of society, society will suffer. So be it: "It is the essence of rights to work as constraints on the pursuit of social goods" (Fried, 1974, page 171).

In fact, demonstration trials have few defenders; most writers consider them inappropriate. In a pointed rejection of the epistemological argument, Hill (1963) wrote

> ... it must be possible ethically to give *every* patient admitted to a trial any of the treatments involved ... If the doctor ... thinks even in the absence of any evidence that for the patient's benefit he ought to give one treatment rather than another, then that patient should not be admitted to the trial. Only if, in his state of ignorance, he believes the treatment given to be a matter of indifference can he accept a random distribution of the patients to the different groups.

Other authors express similar positions in rejecting demonstration trials:

> Sometimes, the initial clinical studies suggest that the new therapy is much more effective than the conventional therapy. In this circumstance, a clinical investigator would not be fulfilling his ethical responsibility to his patients if he planned a randomized comparative trial instead of administering the therapy to consecutive patients. (Gehan and Freireich, 1974.)
>
> Ethically, a physician should only participate in [RCTs] if he or she believes that all treatments under study have potentially equal benefits. (Zelen, 1979.)
>
> Randomization can be applied when the clinician or clinicians in charge of the patient believe that either treatment method, the accepted or the experimental, has essentially equal potential benefit to the patient, and when other factors, such as side effects and personal history, are also of no consequence to the selection process. (Curran, 1979.)

3.6 Experimental Trials and Equipoise

The popular view, expressed in the above quotations, is that demonstration RCTs are ethically impermissible. If the physician believes that A is better for this patient, then he cannot allow ther-

apy to be determined by a random choice between A and B. The same arguments apply even when the physician's belief falls short of absolute conviction. Even if he is only "pretty sure" that A is better, and willing to acknowledge that he might be wrong, the personal care principle precludes his allowing the scientific need for more patients in the B group to lead to his giving B to this patient. The only ethically permissible RCTs are experimental; and even experimental RCTs are prohibited except when the uncertainty about which treatment is better is complete, when the physician is in the state of equipoise, unable to state a preference for either therapy.

This conclusion is the basis for Chalmers' advocacy of "randomizing the first patient" (Shaw and Chalmers, 1970; Chalmers, 1982). He argues that trying a new therapy on a series of patients will usually lead the physician to believe that is better (or worse) than the old, thus destroying his state of equipoise and precluding his participation in an RCT. Now Chalmers views RCTs as "essential for medical progress" (1976, page 137). Thus he urges that the *first* use of new therapy should be in the context of an RCT designed to test its efficacy against the standard therapy.

The requirement of equipoise represents a severe obstacle to starting an RCT. It precludes participation by the proponents of either therapy being tested, restricting the use of this research tool to those sitting squarely on the fence. An even for the physician who has no preference, there are problems with letting a randomizing device choose which therapy a patient receives. Angell (1984) describes one: "... if the physician has no reason to believe that the efficacy of the two treatments is different, then the decision of a fully informed patient will rest on the patient's own preferences... These may be weak or nonexistent in comparing two medicines... but intense when comparing mastectomy with lumpectomy." Equipoise must apply for both physician and patient.

Another consideration is that the physician's state of equipoise must apply to *each* patient: statistical equipoise is not enough. Fried (1974) explains

> I would concede that as to a particular medical condition... viewed across a general population there might be a number of cases where the balance between treatments was equal... But, when a particular patient is involved, with a particular set of symptoms, a particular diagnostic picture, and a particular set of values and preferences, then one may doubt how often a physician carefully going into all of these particularities would conclude that the risks and benefits are truly equal.

3.7 The Problem of Accumulating Evidence

Suppose a physician finds himself in a situation where the perceived cost and benefits of therapies A and B are truly in equipoise with respect to each member of a defined group of patients who themselves have no preferences for one treatment or the other. In this case, he might undertake an RCT to learn (and to show the world) which treatment is better for such patients. Suppose that after applying each therapy to four patients he observes three recoveries in the A group and none in the B group. Now he has some evidence that A is better. It is not strong evidence. It is not "statistically significant" at the conventional 5% level. But it is evidence. And it is sufficient to upset equipoise. If the rate of success under the better treatment is some number k (> 1) times the rate under the one that is worse, and if before this RCT the probability that A is better was one half, then now, given the results for the first eight patients, the probability that A is better is at least $k^3/(1 + k^3)$. If $k = 3$, this is 0.96, if $k = 2$ it is 0.89. If k is only 1.25, the probability is 0.66, which means that the odds that A is the better treatment are at least two to one. It is not yet clear that there is an important difference between the two response rates. But if there is, if either one is much better than the other, then it is now a good bet that the superior treatment is A. The personal care principle requires that the next patient receive not a random choice but A. "... serious consideration of each individual patient's welfare will lead to policies which prevent any clinical trial from producing a clear answer" (Peto et al., 1976). Not all trials present this problem. Some are structured so that all patients are enlisted and all therapeutic decisions are made long before any results can be seen. But, commonly, as the evidence accumulates it upsets equipoise: experimental trials, as they unfold, evolve into demonstration trials. And there is wide agreement that demonstration trials are unethical.

3.8 Concealing the Data

There is a popular solution to this problem, a simple way to prevent experimental trials from evolving into demonstration trials: do not allow those who are conducting the trial to look at the results as they accumulate. That is, to maintain the required state of equipoise, conceal the evidence from the physician until the trial is completed. But there are serious problems with this

solution. Fried (1974, page 35), considering the legal context, writes that

> ... there is a continuing duty on the part of the patient's physician to inform himself about the progress of the experiment and to inform the patient about any significant new information coming out of the experiment that might bear on the patient's choice to remain in the study or to seek other types of therapy.

He points to the relevant legal doctrine and concludes

> Nor would the device, by which only a supervising committee and not the patient's physician has access to the results of the experiment for a determined period of time, insulate the physician from the consequences of this doctrine.

And Marquis (1983), in an ethical analysis of the argument for concealing the data, observes that

> This is quite a remarkable argument ... [P]hysicians who would not enroll a patient in a study on moral grounds should be prevented from doing so by withholding from them information needed to make the moral judgment?

3.9 Adaptive Trials

Another response to the problem represented by accumulating evidence is to depart from 50-50 randomization. As the evidence that A is better grows, incoming patients are still randomized to A or B, but the probability of B is reduced. The arguments for such procedures are usually utilitarian: they reduce the number of patients who receive the disfavored treatment, and, compared to 50-50 randomization, improve participants' chances of getting the better treatment. But the ethical problems are clear: after finding enough evidence favoring A to require reducing the probability of B, the physician obeying the personal care principle must see that the next patient gets A, not just with high probability, but with certainty.

3.10 Informed Consent

Perhaps the ethical problems with RCTs can be avoided by proper use of informed consent. If the patient, when fully informed about the situation, voluntarily chooses to give up his right to personal care and agrees to allow his treatment to be selected by chance, then that also is his right. The physician might explain that a new treatment is being tested against a conventional one, and that he personally thinks the new one will prove to be better (or worse), although it is far from certain. He might also point out that some colleagues whose opinions he respects prefer the other treatment. Under such conditions, many patients might be quite willing to participate in an RCT. They might behave altruistically, releasing the physician from his obligation to do what in his judgment is best for them, and even consenting to the concealment of the accumulated evidence.

> Provided no law is violated, our mores sanction almost any experiment if participation is voluntary, after full disclosure of its implications. (Zeisel, 1970.)

Although proper use of informed consent can allow many randomized clinical trials to proceed, it is no panacea. Some see ethical or legal problems with *asking* for consent from a person who is seriously ill and who has come to the physician for help, not to volunteer as a research subject (Curran, 1979; Marquis, 1983). And it is not clear that anyone has the right to give consent for infants and others who are incapable of autonomous decision-making to participate in a study where therapy is selected by any criterion other than the personal care principle.

A more prosaic reason why informed consent is not a panacea is that in many cases it simply won't work. Sometimes too few patients will agree to participate in the study. And sometimes the consent process is simply too uncomfortable and time-consuming. Many physicians are reluctant to share their uncertainties with patients, and many patients do not want to hear about such uncertainties. Also, many patients know so little about research that to make them truly understand what they are being asked to do and why, to make their consent truly "informed," would require more time and effort than can reasonably be invested.

3.11 The Dilemma

The basic dilemma that plagues randomized clinical trials occurs because the physician-investigator is playing two roles. As a clinician he is bound by the personal care principle to make the therapeutic decision solely on the basis of what, in his professional judgment, is best for this individual patient, while as a scientist he needs to assign adequate numbers of patients to each treatment and to let the assignments be made by a randomizing device, so that his judgment is not allowed to bias the trial.

Although some writers emphasize both the possibility that the physician's treatment preference might be wrong and the great social value of an RCT that convinces clinicians that a standard ther-

apy is in fact inferior (e.g., Gilbert, McPeek and Mosteller, 1977; Silverman, 1980), most insist that the personal care principle is paramount. Angell (1984) writes

> Physicians traditionally act in the best interests of each patient under their care, and patients expect this of their physician. If this commitment to the patient is attenuated, even for so good a cause as benefits to future patients, the implicit assumptions of the doctor-patient relationships are violated. I have no doubt that we would lose more than we would gain by adopting such an approach.

This is the position promulgated by the World Medical Association in the Declaration of Helsinki (1964): "Concern for the interests of the subject must always prevail over the interests of science and society." Hill (1963) put it even more simply: "... the ethical obligation always and entirely outweighs the experimental."

4. A RANDOMIZED CLINICAL TRIAL

In 1985 a group at a major neonatal intensive care center published the results of a randomized clinical trade of two therapies for newborn infants with severe respiratory failure (Bartlett et al., 1985). This center and others had experienced very high mortality rates among such infants when conventional therapy was applied, and some of these physicians had been leaders in the development of an alternative treatment based on extracorporeal membrane oxygenation, or ECMO. They had previously published the results of an uncontrolled study in which ECMO had been used on 55 "moribund" infants. Among the 15 infants whose birth weight was less than 2 kg 3 survived, while among the 40 whose birth weight was more than 2 kg, 28 (70%) survived. With this experience behind them, they designed and carried out an RCT to compare ECMO with conventional therapy.

The published report makes it clear that this was not an experimental but a demonstration trial. The participating physicians had a definite treatment preference: "We anticipated that most ECMO patients would survive and most control patients would die... ." Nevertheless, "... we were compelled to conduct a prospective randomized study..." in which critically ill infants were assigned to ECMO or control on the basis of a random draw. This is particularly disturbing to me as a statistician because it is we statisticians who are largely responsible for creating the attitudes and assumptions that "compelled" this study. This is not to deny that some physicians are among the most enthusiastic proponents of the doctrine that nothing is scientifically established until it has been shown in a properly controlled randomized clinical trial, so that those who would resolve a therapeutic controversy are "compelled" to do an RCT. But because this doctrine addresses fundamental questions of experimental design and inference, it is within the purview of statistics and is the responsibility of statisticians, from whom the physicians learned it. If Freireich (1979) is correct in his assessment that "... the brilliant success of the RCT has now become a form of intellectual tyranny," statisticians should help to redress matters. In the next section I propose that the above doctrine springs at least in part from adherence to a faulty statistical principle.

These investigators saw themselves in a "... scientific/ethical dilemma" in which they felt compelled to conduct an RCT "but reluctant to withhold a lifesaving treatment from alternate patients simply to meet conventional random assignment technique" (Bartlett et al., 1985). Their approach to this dilemma was to use an adaptive randomization scheme, one of the "randomized play-the-winner" treatment assignment rules of Wei and Durham (1978). This rule promised that, if things went as expected, the fraction of patients from whom the lifesaving treatment would be withheld would probably, but not surely, be considerably less than the full 50% required under "conventional random assignment."

The rule used was equivalent to the following procedure: The treatments are coded A and B. Initially there are two balls in an urn, one marked A and one B. Each patient's treatment is determined by drawing a ball from the urn. After each draw the indicated treatment is applied, the result is observed, the ball is put back in the urn, and a new ball is added. If the treatment was successful the new ball carries the same treatment letter (A or B), and if not, it carries the other letter. The trial ends when 10 balls of one type have been added, with that treatment then being declared the winner. This scheme requires at least 10 and at most 19 patients in the study, with the actual number being a random variable. How many of these patients actually will have received the treatment that is declared the loser is also a random variable, with possible values ranging from "none" to "all." Under the assumption that one treatment provides a substantially greater chance of survival than the other, the probability is very high that this randomized play-the-winner rule will select as the winner the treatment that is actually better. If, for example, the survival rate under the better treat-

ment is at last 0.8 and if it is at least 0.4 greater than the rate under the worse one, then the probability that the better treatment will be declared the winner is at least 0.95. and "For the probabilities actually thought to hold, namely $P_A = 0.9$ for ECMO and $P_B = 0.1$ for conventional therapy, the probability of selection of the better treatment after 10 patients not only exceeded 0.95 but equalled 1.0" (Bartlett et al., 1985).

As it turned out, the first patient was assigned ECMO and survived, the second was assigned conventional therapy and died. The next eight patients were assigned ECMO. All survived. At this point 10 ECMO balls had been added to the urn, so ECMO was declared the winner. Then two more patients were assigned ECMO, and they both survived. The authors concluded that "... this study proves that ECMO improves survival when compared to conventional treatment."

The adaptive randomization scheme had been successful in assigning to most patients the treatment that proved to be the winner. Only one patient had been given the losing therapy. Levine (1986, page 210) cites this result with approval: "Consider what the death rate might have been in a conventional RCT." However, the authors expressed not satisfaction but disappointment with this aspect of their study: "In retrospect it would have been better to begin with two or three pairs of balls, which probably would have resulted in more than one control patient." In this, they were correctly anticipating the criticism made by Ware and Epstein (1985), who argued

> ... the results are not completely convincing. Why not? Because only one patient received the standard therapy, so that the interpretation of the study depends strongly on the belief that eligible patients would have experienced poor survival in the absence of [ECMO]

and concluded that

> Further randomized clinical trials using concurrent controls and addressing the ethical aspects of consent, randomization, and optimal care will be difficult but remain necessary.

The question of whether the parents' informed consent makes it ethically acceptable to include a critically ill infant in a study such as this is difficult. With respect to this particular study, the question is hypothetical, because the parents of the infant who was randomized to conventional therapy were *not* informed about the trial and its rationale, about their physician's judgment about the two available therapies, about the fact that conventional therapy was selected for their child, not on the basis of that judgment, but by a random device, and, most important, about their right not to participate in the study, not to accept the randomly selected therapy. For this, too statisticians must share the responsibility. The investigators were using Zelen's (1979) "randomized consent" design.

The assumption behind the randomized consent design is that there is no need to inform patients assigned to "standard therapy" about the trial (and about the fact that the therapeutic choice was made in their case by a randomizing device) and no need to ask for their consent to participate: "... every patient expects to receive the best standard therapy. It is only when there are departures from this expectation that the patient should be so informed" (Zelen, 1979). In the present application, this argument is particularly disingenuous in view of the fact that the study was done in a center where "Many [of the babies] were referred from other level III intensive care centers *because of the availability of ECMO in this institution*" (Bartlett et al., 1985; italics added).

Science desires randomized clinical trials, it does not demand them. Writers promoting wider use of randomized trials often dramatize their merits by comparing them to trials that use historical controls (e.g., Byar, 1979). Randomized trials *are* often better in many important ways than historical control trials. But both types of trial can produce misleading results and *neither can prove anything*. Historical controls require more cautious interpretation, but this is only a matter of degree.

> ... the difference between the RCT and the observational, retrospective study is not the difference between good and bad science, truth and falsity, but a difference between varying degrees of confidence. (Fried, 1974, page 159.)

Moreover the importance of randomized trials is exaggerated by the false dichotomy that is implied when historical control trials are the only alternative discussed. In fact, many of the weaknesses of historical controls can be avoided by using concurrent nonrandomized controls. And of course comparing a treatment with both historical and concurrent controls can be provide even stronger evidence about its efficacy.

In the case of the ECMO trial, a multicenter observational study might have been done in which patients at centers such as the University of Michigan, where ECMO was favored, would be compared to the patients at centers such as Johns Hopkins where only "conventional" therapy was available. The protocol could have specified rigid criteria for

inclusion in the study, for diagnosis and evaluation, and for exactly what data were to be recorded at entry and subsequently, with precise definitions to ensure that the data would be recorded uniformly for all patients. Those responsible for monitoring the study could have seen that all of the centers produced high quality data, with specified procedures for detecting errors and preventing missing observations. (Here I am paraphrasing some of Byar's, 1979, description of a "properly run" randomized trial.) Such a multicenter study would have had some scientific disadvantages compared to a comparable multicenter randomized study, but it would also have had the important advantage that every baby in the study would be receiving what his physician considered to be the best care that he could deliver.

5. THE RANDOMIZATION PRINCIPLE

R. A. Fisher is generally credited with introducing deliberate randomization into experimental design. In *Design of Experiments* (1935; Section 20), he gave an example showing how a test of significance could be calculated on the basis of the probability distribution that is created when treatments are assigned to experimental units at random. He also asserted (Section 21) that "... the physical act of randomization ... is necessary for the validity of any test of significance ... ." thereby planting the seed of the doctrine that lives today as the randomization principle. This principle asserts that deliberate randomization creates the only probability distributions on which valid statistical inferences can be based (Royall, 1976, 1983). As Bearman put it "... the only logical bases that we know ... for making generalizations from data all rest on the foundation of randomization." Thus "If any of you need my help, or that of any of my biostatistical colleagues, random assignment of patients will be a *sine qua non* ... " (Bearman, 1976).

The randomization principle is one important factor contributing to the conviction that RCTs are not just desirable but also a scientific necessity. This is because the principle implies that statistically valid inferences from nonrandomized studies are impossible. Hence Silverman's (1980) observation that in a particular study "A declaration of statistical significance is meaningless ... " because "The key requirement of Fisherian design was not fulfilled: random assignment of treatments ... " and Cowan's (1981) assertion, quoted in Section 1, that nothing can be learned from nonrandomized trials.

The randomization principle has been accepted by some prominent statisticians (e.g., Yates, 1948; Stuart, 1962). It has been rejected by others (e.g., Neyman and Pearson, 1937; Savage, 1962). Cornfield (1959), using the term "experiment" to refer to randomized studies, wrote: "The proposition that some inherent logical incompetence attaches to an inference based on observational [nonrandomized], as distinguished from experimental evidence seems to have little to commend it beyond the great positiveness with which it is sometimes asserted." And Fisher (1960) himself backed down, adding a new section (Section 21.1) to the seventh edition of his book, in which he claimed that his original example "... was in no sense put forward to supersede the common and expeditious tests ... " that are based on probability models other than the one created by deliberate randomization, such as the normal or "Gaussian" distribution models.

Statistical theory explains why the randomization principle is unacceptable. It does this in terms of the concepts of conditionality (ancillarity) and likelihood (Berger and Wolpert, 1988; Cox and Hinkley, 1974; Royall, 1976). The conditionality principle asserts that when an ancillary statistic C is present, inferences should be based on the conditional probability model given the observed value of C. The problem this creates for the randomization principle is that the statistic representing the result of the randomization is ancillary; thus the conditional randomization distribution is degenerate, assigning probability one to the allocation pattern that was actually used. The only "inference" from the observed data that this conditional distribution supports is "I saw what I saw" (Royall and Cumberland, 1981, Section 2.3 and Rejoinder).

6. CONCLUSION

"Randomized clinical trials remain the most reliable method for evaluating the efficacy of therapies" (Byar et al., 1976). This does not imply that studies that use historical or nonrandomized concurrent controls have little or no scientific value. We can learn without randomization. Epidemiologists routinely find it impossible to obtain randomized control groups. Yet they have made findings of enormous importance to science and health. In this context, it is noteworthy (and perhaps statistically significant) that those who were convinced by the observational evidence that cigarette smoking increases the risk of lung cancer, and who urged that the population should be so warned, were persistently and caustically criticized by Fisher, the father of the randomization principle, who railed against "... the fallacious conclusions ... about the danger of cigarettes" (Fisher, 1959).

In some situations clinical investigations are badly needed, but nonrandomized controls are the

only ones that can be obtained ethically. Statisticians should be more sensitive to this point when they serve as designers, reviewers and critics of clinical studies. Although it is important that we understand, teach and exploit the advantages of RCTs, it is also important that we not exaggerate the disadvantages of carefully conducted and analyzed studies that use nonrandomized controls: "...we should not proceed on the fallacious assumption that where there is no randomization there is no truth" (Fried, 1974, page 160).

REFERENCES

ANGELL, M. (1984). Patients' preferences in randomized clinical trials. *N. Engl. J. Med.* **310** 1375-1387.

BARTLETT, R. H., ROLOFF, D. W., CORNELL, R. G., ANDREWS, A. F., DILLON, P. W. and ZWISCHENBERGER, J. B. (1985). Extracorporeal circulation in neonatal respiratory failure: A prospective randomized study. *Pediatrics* **76** 476-487.

BEARMAN, J. E. (1976). Biostatistical principles in dermatopharmacology. *J. Investigative Dermatology* **67** 679-681.

BERGER, J. O., and WOLPERT, R. L. (1988). *The Likelihood Principle*, 2nd ed. IMS, Hayward, Calif.

BYAR, D. P. (1979). The necessity and justification of randomized clinical trials. In *Controversies in Cancer: Design of Trials and Treatment* (H. J. Tagmon and M. J. Staquet, eds.) 75-82. Masson, New York.

BYAR, D. P., SIMON, R. M., FRIEDEWALD, W. T., SCHLESSELMAN, J. J., DEMETS, D. L., ELLENBERG, J. H., GAIL, M. H. and WARE, J. H. (1976). Randomized clinical trials: Perspective on some recent ideas. *N. Engl. J. Med.* **295** 74-80.

CHALMERS, T. J. (1976). Contribution to discussion: How to turn off an experiment. In *Ethical Safeguards in Research on Humans* (J. D. Cooper, ed.) 119-143. Interdisciplinary Communication Associates, Washington, D.C.

CHALMERS, T. J. (1982). A potpourri of RCT topics. *Controlled Clinical Trials* **3** 285-298.

CHAPMAN, C. B. (1984). *Physicians, Law, and Ethics*. New York Univ. Press.

CORNFIELD, J. (1959). Principles of research. *Amer. J. Mental Deficiency* **64** 240-252.

COWAN, D. H. (1981). The ethics of trials of ineffective therapy. *IRB: A Review of Human Subjects Research* **3** 10-11.

COX, D. R. and HINKLEY, D. V. (1974). *Theoretical Statistics*. Chapman and Hall, London.

CURRAN, W. J. (1979). Reasonableness and randomization in clinical trials: Fundamental law and governmental regulations. *N. Engl. J. Med.* **300** 1273-1275.

FISHER, R. A. (1935). *Design of Experiments*. Oliver and Boyd, Edinburgh.

FISHER, R. A. (1959). *Smoking: The Cancer Controversy*. Oliver and Boyd, Edinburgh.

FISHER, R. A. (1960). *Design of Experiments*, 7th ed. Oliver and Boyd, Edinburgh.

FREEDMAN, B. (1987). Equipoise and the ethics of clinical research. *N. Engl. J. Med.* **317** 141-145.

FREIREICH, E. J. (1979). Invited remarks on: Levine, R. J. and Lebacqz, K. L. Some ethical considerations in clinical trials. *Clinical Pharmacology and Therapeutics* **25** 728-746.

FREUND, M. E. (1982). Surgical Research. In *Human Subjects Research* (R. A. Greenwald, M. K. Ryan and J. E. Mulvihill, eds.) 169-179. Plenum Press, New York.

FRIED, C. (1974). *Medical Experimentation: Personal Integrity and Social Policy*. North-Holland, Amsterdam.

GEHAN, E. A. and FREIREICH, E. J. (1974). Non-randomized controls in cancer clinical trials. *N. Engl. J. Med.* **290** 198-203.

GILBERT, J. P., MCPEEK, B. and MOSTELLER, F. (1977). Statistics and ethics in surgery and anesthesia. *Science* **198** 684-689.

HILL, A. B. (1963). Medical ethics and controlled trials. *Brit. Med. J.* **1** 1043-1049.

LEVINE, R. J. (1986). *Ethics and Regulation of Clinical Research*, 2nd ed. Urban and Schwarzenberg, Baltimore.

MARQUIS, D. (1983). Leaving therapy to chance. *Hastings Center Report* **13** 40-47.

MEDICAL RESEARCH COUNCIL (1964). Responsibility in investigations on human subjects. *Brit. Med. J.* **2** 178-180.

MULDER, J. H. (1981). The ethics of clinical trials of ineffective therapy. *IRB: Review of Human Subjects Research* **3** 9-10.

NEYMAN, J. and PEARSON, E. S. (1937). Note on some points in "Student's" paper on "Comparison between balanced and random assignments of field plots." *Biometrika* **29** 380-388.

PETO, R., PIKE, M. C., ARMITAGE, P., BRESLOW, N. E., COX, D. R., HOWARD, S. V., MANTEL, N., MCPHERSON, K., PETO, J. and SMITH, P. G. (1976). Design and analysis of randomized clinical trials requiring prolonged observation of each patient: Part 1, Introduction and design. *Brit. J. Cancer* **34** 585-612.

ROYALL, R. M. (1976). Current advances in sampling theory: Implications for human observational studies. *Amer. J. Epidemiology* **104** 463-474.

ROYALL, R. M. (1983). Comment on "An evaluation of model-dependent and probability-sampling inferences in sample surveys" by M. H. Hansen, W. G. Madow and B. J. Tepping. *J. Amer. Statist. Assoc.* **78** 794-796.

ROYALL, R. M. and CUMBERLAND, W. G. (1981). An empirical study of the ratio estimator and estimators of its variance (with discussion). *J. Amer. Statist. Assoc.* **76** 66-88.

SAVAGE, L. J. (1962). *The Foundations of Statistical Inference*. Methuen and Company, London.

SCHAFER, A. (1982). The ethics of the randomized clinical trial. *N. Engl. J. Med.* **307** 719-724.

SHAW, L. W. and CHALMERS, T. C. (1970). Ethics in cooperative clinical trials. *Ann. New York Acad. Sci.* **169** 487-495.

SILVERMAN, W. A. (1980). *Retrolental Fibroplasia: A Modern Parable*. Grune and Stratton, New York.

STUART, A. (1962). *Basic Ideas of Scientific Sampling*. Hafner, New York.

TUKEY, J. W. (1977). Some thoughts on clinical trials, especially problems of multiplicity. *Science* **198** 679-684.

VEATCH, R. M. (1981). *A Theory of Medical Ethics*. Basic Books, New York.

WARE, J. H. and EPSTEIN, M. F. (1985). Extracorporeal circulation in neonatal respiratory failure: A prospective randomized study. *Pediatrics* **76** 849-851.

WEI, L. J. and DURHAM, S. (1978). The randomized play-the-winner rule in medical trials. *J. Amer. Statist. Assoc.* **73** 830-843.

WORLD MEDICAL ASSOCIATION (1948). Declaration of Geneva. Reprinted in Beauchamp, T. L. and Childress, J. F. (1983). *Principles of Biomedical Ethics*. Oxford Univ. Press, New York.

YATES, F. (1948). Comment on "The validity of comparative experiments" by F. J. Anscombe. *J. Roy. Statist. Soc. Ser. A* **51** 181-211.

ZEISEL, H. (1970). Reducing the hazards of human experiments through modifications in research design. *Ann. New York Acad. Sci.* **169** 475-486.

ZELEN, M. (1979). A new design for randomized clinical trials. *N. Engl. J. Med.* **300** 1242-1246.