Rejoinder

James H. Ware

INTRODUCTION

The discussants have raised many important issues about the ECMO study and about clinical research in general. The diversity of views expressed about our study is not surprising, given the difficult issues that it raises. It is especially interesting that some of the discussants believe that no patients should have been randomized, while others think that randomization was stopped to quickly.

I do believe that some of Dr. Berry's comments do not address scientific issues, especially those that question the motivation of the investigators, suggest that we could not think independently about the strengths and weaknesses of alternative designs, or suggest that we would have knowingly given an inferior therapy to study patients. I will not address these comments. Instead, my reply focuses on the following points: 1) the role of randomization in medical research, 2) the ethical justification for randomization in the ECMO study, 3) randomized consent, 4) study design and 5) data analysis. Finally, I will comment on some of the lessons to be learned from this study and its reception.

THE ROLE OF RANDOMIZATION IN MEDICAL RESEARCH

Dr. Berry is very critical of the role of randomization in therapeutic research. He argues that it is both unethical and unnecessary; unethical because we always have some preference among treatments, and unnecessary because data banks and registry studies can provide equally valid information about the relative efficacy of therapies. Although others have taken Dr. Berry's position, most scientists now recognize the unique role of randomization in the conduct of experiments and clinical trials. It would not be useful to repeat the arguments in favor of randomization here (see, for example, Byar et al., 1976), but a few points deserve emphasis.

Biostatisticians and physicians involved in therapeutic research are accustomed to hearing reports of new therapies that give outstanding results in uncontrolled studies. Regrettably, only a small fraction of these new treatments prove to be superior to standard therapies when evaluated in a randomized trial. A distinguished expert in cancer clinical trials, Dr. Charles Moertel, spoke humorously to this point as keynote speaker at the 1980 meeting of the Society for Clinical Trials in a talk entitled "How to succeed in clinical trials without really trying." The key to the strategy was the use of nonrandomized studies to generate very favorable results. Similarly, early uncontrolled studies of coronary bypass surgery led to exaggerated claims about its efficacy in the treatment of coronary artery disease (Detre, 1984). Randomized trials showed that some patient groups benefited from bypass surgery, while others did as well with medical management. Dr. Moertel's message would apply equally well to research on many other medical conditions. It is important to be skeptical about 'breakthroughs' and to examine them carefully.

The limitations of observational studies arise from the natural heterogeneity of human response. Although predictors of patient outcome have been identified in many settings, these predictors usually explain a small part of the variability among patients or study participants. The Framingham Heart Study, for example, has played a central role in identifying several important risk factors for myocardial infarction (MI), but the high risk group identified by these risk factors has a 2-year probability of MI of only about .1 (Shurtleff, 1974). Our limited ability to predict outcome using quantitative methods creates the potential for bias in subjective methods of assigning patients to therapy. The high success rates sometimes reported from nonrandomized studies in high-risk diseases suggest that clinicians can identify patients with a favorable prognosis even when they cannot quantify their criteria.

It is important to recognize that Bayesians and frequentists agree about the purpose and value of randomization. Rubin (1978), for example, shows that, when the mechanisms that sample experimental units, assign treatments, and record data are not ignorable, the Bayesian must model them. The resulting inferences then become sensitive to model specification. Randomized trials can, of course, give misleading results, but randomization provides a framework for calculating error rates or posterior probabilities that do not depend on assumptions about comparability.

Dr. Berry argues that therapies could be evaluated by analyzing registry data gathered in large national data banks. This is unrealistic. Registries that do not include well-defined entry criteria, complete coverage of study populations, careful quality control procedures, and other features of careful research are of little value. Carefully designed and implemented registry studies have proven to be of value, for example in the study of coronary artery bypass surgery (Detre, 1984) and percutaneous transluminal coronary

angioplasty (Kent et al., 1982), but these studies are costly and, in some ways, more difficult than randomized trials. Even in these carefully managed studies, little is known about the potential for bias. Some scientists and advocates are promoting wider use of registries in medical research, but it would be a tragic setback to medical science if registries replaced randomized trials as the mainstay of therapeutic research.

Given that randomization has an important role in assuring the validity of clinical trials, one must decide when randomization is ethically justified. Drs. Kass and Greenhouse note in quoting Sir A. Bradford Hill that randomized trials cannot be used when "treatment cannot ethically be withheld." It can be difficult to determine when this is so, and medical and statistical observers can have widely varying opinions about the same situations, as the discussants have illustrated. We turn now to the ethics of randomization in the EMCO study.

WAS A RANDOMIZED TRIAL ETHICALLY JUSTIFIABLE?

Our study design attempts to address the potential conflict between two goals, that of treating individual patients optimally and that of carefully evaluating a new therapy. Dr. Royall argues that it was unethical to begin a randomized trial in 1986, while Dr. Begg argues that randomization should have been continued after the first 19 patients were enrolled. This is not surprising. One's view of the ethics of our study depends directly on one's assessment of the evidence available to us.

Dr. Royall's position seems to me to be internally inconsistent. He may be confident that *he* would not have been prepared to begin a randomized trial in 1986 (though even this assertion is easier to make now that our study has been completed). He would acknowledge, however, that different investigators will draw different inferences from the same data. Moreover, he does not have full access to either the data or the expertise that my medical colleagues and I drew on in evaluating ECMO. As Dr. Kass and others have said in public discussion of this issue, it is not appropriate for us as statisticians to conclude from the clinical trial data alone that ECMO is or is not superior.

To emphasize the key point, we were not convinced by the limited evidence available in 1986 that ECMO was superior to or even as effective as optimal medical therapy. Readers should note that some of the data cited by discussants were not available at that time. Even as recently as 1988, a report of national experience on the use of ECMO for a variety of conditions listed 715 ECMO treated patients (Toomasian et al., 1988). Of these, about 100 were seen by the end of 1985, of whom roughly 50% had a primary diagnosis of PPHN. In 1986, the Harvard hospitals were the only hospitals in New England that offered ECMO. ECMO is highly invasive, and therefore both expensive and potentially dangerous. Harvard physicians were especially concerned about the potential for serious neurological complications resulting from brain hemorrhages caused by prolonged ECMO treatment.

Even now, medical opinion about the efficacy of ECMO is far from unanimous. After 10 patients had been entered into phase 2 of our study, we asked four colleagues, two physicians with expertise in neonatology and two experienced biostatisticians, to review the progress of the study. They advised us to resume randomization, arguing that the possibility of long-term morbidity remained and that our mortality data were not definitive. We understood this perspective, but were not willing to resume randomization.

Dr. Royall makes the interesting observation that the prior distributions proposed in my paper can be combined with the data collected in the Michigan study to obtain posterior probabilities not very different from those we calculated from our first 19 patients. This is not an artifact of the prior. The likelihood based on the 10 patients in the Michigan study is concentrated above the 45° line. Dr. Royall concludes from this calculation that it was unethical to begin our study. We agree that the Michigan study was an important piece of evidence, but believe that treatment strategies for hundreds if not thousands of future patients should not be determined from such scanty evidence.

In summary, the superiority of ECMO was far from clear when we began our study, despite optimistic claims and our own hopes. Dr. Royall suggests that a nonrandomized trial, for example a trial comparing ECMO and CMT at different institutions, might have been preferable. In the same vein, Drs. Kass and Greenhouse note the need for better methods for using historical information, and for the design, conduct, and reporting of nonrandomized studies. Nonrandomized studies could be of value in studies of therapies offering potential for great benefit. Standards could be based on the already well-developed criteria for excellence of epidemiologic research.

RANDOMIZED CONSENT

The method of randomized consent is a new approach to randomization. It raises subtle ethical and scientific issues that should be carefully explored by medical investigators. One clarification is important, however, in responding to the concerns of Drs. Berry and Royall. In our trial, the use of randomized consent did not result in denial of ECMO therapy to study patients. The physicians involved in the ECMO study

had considered whether ECMO would be used either in routine treatment of PPHN or for rescue therapy in medically treated infants near death. They concluded that they were not prepared to use ECMO in either way without additional evidence of its efficacy. Thus, during the investigation, ECMO could be given as treatment for PPHN only in the context of the clinical trial.

This issue is connected to a broader discussion of the availability of treatment that has been stimulated by the AIDS epidemic. The usual practice in medical research, based on guidelines established by the U.S. Government and followed by most hospital Institutional Review Boards, is to give experimental therapies only in the context of a research study. In May 1987, however, the Food and Drug Administration issued new regulations permitting the use of experimental drugs in the treatment of 'desperately ill' patients. These regulations may make it impossible to conduct standard randomized clinical trials in some settings. If so, statisticians will be asked to develop alternative designs and strategies for data analysis.

Given that ECMO was not available to nonstudy patients, the question of whether to discuss the study with patients assigned to CMT becomes more subtle. We felt that discussion of the trial, either before or after randomization, would add to the stress experienced by parents of very sick infants without providing a corresponding benefit. This policy was compassionate, but inconsistent with another important goal of medical care, openness with patients and families about treatment decisions. Perhaps medical ethicists can help us examine this issue more carefully.

STUDY DESIGN

Our study design was an attempt to obtain useful comparative data on ECMO and CMT in a setting where early results might show a strong trend favoring one therapy. This situation has a strongly Bayesian flavor. Kadane et al. (1980) and Kadane (1986) discuss Bayesian approaches to randomized clinical trials. As Efron (1986) notes, however, these methods are difficult to implement and explain to nonspecialists. For these reasons, and because the medical literature has a strong frequentist orientation, I chose a design with a frequentist interpretation.

When a clinical trial is designed to have good power against a specified alternative, the design has a decision-theoretic interpretation. It provides a strategy for using study data to choose between two states of nature, with small probabilities of making the wrong choice. The selection paradigm used in the Michigan study implied that either therapy had a 50% chance of being selected if the therapies were equally effective. This approach would lead to frequent acceptance of new therapies that do not improve treatment.

Several discussants point out that we might have reached the end of the randomized phase with nearly comparable results in the two treatment groups. In fact, our single center trial was constrained to be small. Thus, the study had very low power against more modest treatment differences. Drs. Armitage and Coad, among others, have, however, made an important point in noting that the two-stage approach used in the ECMO study could be used with any sequential procedure. Designs based on the likelihood ratio statistic, for example, might be more generalizable and more robust to unanticipated outcomes of the study. Dr. Hardwick makes several important observations about the difficulties that arise when adaptive designs are used in medical research.

Although Drs. Kass and Greenhouse find the log odds attractive, it is not always an appropriate parameter for medical decision making. Here, I believe that the medical significance of the evidence is more closely related to the difference in survival rates.

DATA ANALYSIS

Our data analysis was based on the assumption that patients from phases 1 and 2 could be combined. Drs. Armitage and Coad correctly note that temporal drift could produce bias in two-stage designs. In this instance, the data suggest that such temporal drift as exists biases the study against ECMO.

Dr. Berry comments that we "couldn't make a Neyman-Pearson inference." This is not entirely correct. Our paper reports a valid frequentist test of significance. Because the study was truncated, we obtained an upper bound for the P value. This upper bound could have been translated into an exact lower bound for the 95% confidence interval for the risk ratio. This lower bound would have been near 1. Instead, we used profile likelihood methods to calculate a 95% confidence interval. Drs. Lin and Wei report that the profile-likelihood confidence interval has limited accuracy with small sample sizes and calculate a conservative one-sided 95% confidence interval based on exact methods. Their lower bound is 0.035. This is an improvement over our result, for which I thank Drs. Lin and Wei.

Several discussants misunderstood my purpose in reporting a Bayesian analysis of the first 19 randomized cases. My intent was to show that the phase 1 data provided strong evidence that ECMO is at least as effective as CMT under a range of plausible prior beliefs about relative efficacy and thus made it difficult to justify further randomization. Drs. Kass and Greenhouse examine the evidence under a much wider family of priors and come to very similar conclusions.

I still prefer, however, to calculate the probability that ECMO is inferior, given the data in hand. This is the probability that I will choose an inferior treatment if I stop the trial and choose ECMO.

CONCLUSIONS

The ECMO study brings many of the ethical and scientific issues present in most clinical trials into sharp relief. For example, the tension between optimal treatment and rigorous comparative study often leads to disagreements about study management. I do not believe that ethical care requires that each physician be required to choose the therapy that he or she considers best when there is little or no evidence at hand to guide this choice. This approach would make it difficult, if not impossible, to compare treatments. When the relative efficacy of two or more treatment strategies is in doubt, the randomized clinical trial provides a method for pursuing the larger ethical goal of increasing medical knowledge, to the ultimate betterment of the community, without compromising the care of the individual patient. When evidence begins to accumulate favoring one therapy, however, this fragile equilibrium is threatened. Concepts such as clinical equipoise address but do not fully resolve this issue. This issue was brought into sharp focus in the ECMO study. Perhaps further research or imaginative applications of methods already in hand can provide additional guidance on strategies for delivering optimal medical care to the individual patient while continuing the process of careful evaluation of new therapies.

I have learned more about several aspects of clinical trials research from this study. First, my belief in the importance of randomization in most therapeutic studies has been strengthened. The ECMO experience has, however, raised the question of the potential value of nonrandomized strategies in settings where therapies may have very different efficacies. Further research and experience is needed to determine how successful this approach can be.

In retrospect, the essential idea in our design was very simple. We planned to randomize patients until the data became strongly suggestive, but not definitive, and then to continue with a nonrandomized study of the more promising therapy. Thus, our design is a variant on the use of historical controls. Drs. Armitage and Coad have pointed out that this idea could be applied to other designs, including those based on sequential procedures such as the SPRT. This idea deserves further study. I hope that readers will be inspired by this experience and by the insightful comments of the discussants to explore other methods for sharing the advances obtained through therapeutic research with both current and future patients.

ADDITIONAL REFERENCES

BYAR, D. P., SIMON, R. M., FRIEDEWALD, W. T., DEMETS, D. L., ELLENBERG, J. J., GAIL, M. H. and WARE, J. H. (1976). Randomized clinical trials: Perspectives on some recent ideas. *New England J. Med.* **295** 74–80.

DETRE, K. (1984). Nonrandomized studies of coronary artery bypass surgery. *Statist. in Medicine* **3** 389–398.

EFRON, B. (1986). Why isn't everyone a Bayesian? (with discussion). *Amer. Statist.* **40** 1–11.

KADANE, J. B. (1986). Progress toward a more ethical method for clinical trials. *J. Med. Philos.* **11** 385–404.

KADANE, J. B., DICKEY, J. M., WINKLER, R. L., SMITH, W. S. and PETERS, S. C. (1980). Interactive elicitation of opinion for a normal linear model. *J. Amer. Statist. Assoc.* **75** 845–854.

KENT, K. M., BENTIVOGLIO, L. G., MYLER, R. K., SIMPSON, J., STERTZER, J. et al. (1982). Percutaneous transluminal coronary angioplasty: Report from the Registry of the NHLBI. *Amer. J. Cardiol.* **49** 2011–2020.

RUBIN, D. B. (1978). Bayesian inference for causal effects: The role of randomization. *Ann. Statist.* **6** 34–58.

SHURTLEFF, D. (1974). *The Framingham Heart Study: An Epidemiologic Investigation of Cardiovascular Disease* (W. B. Kannel and T. Gordon, eds.) **30,** DHEW Publication No. 74-599. NIH, Bethesda, Md.

TOOMASIAN, J. M., SNEDECOR, S. M., CORNELL, R. G., CILLEY, R. E. and BARTLETT, R. H. (1988). National experience with extracorporeal membrane oxygenation for newborn respiratory failure: Data from 715 cases. *ASAIO Trans.* **34** 140–147.