Comment

Robert H. Bartlett and Richard G. Cornell

Richard Royall argues that the prospective randomized clinical trial is a sacred cow which achieved undeserved popularity if not reverence among the scientific medical community. He acknowledges that statisticians, by expounding the merits of the randomized trial, have contributed to the general perception that a drug, technique or device has not been adequately tested until it has been subjected to a prospective randomized clinical trial. He argues that comparison to historical controls, good results with large numbers of patients with a new treatment and simple common sense can provide more than adequate evidence that a new treatment is better than the standard treatment. Moreover if the preliminary information leading to the randomized study indicates that the new treatment will be better than the old, Royall states that it is unethical to even conduct a randomized trial, based on the principle of personal care. That is, he assumes that it is the moral duty of any physician to use whatever treatment the physician considers to be the best treatment.

We agree with many of the points made by Royall. Since our study was published, many experts have come forward in the scientific press and the lay press pointing out that randomized studies of life support systems pose such difficult ethical problems that randomized studies may not be justified. Ware and Epstein (1985) were articulate critics of our attempt to deal with both the ethical and statistical issues by using an adaptive design. Yet when an additional randomized study of ECMO in neonatal respiratory failure was carried out by them at Harvard, an adaptive design was used (O'Rourke et al., 1989). The reason for this choice (Ware, 1989) was the same as ours, namely, the desire to minimize the number of patients on standard therapy if, in fact, ECMO is superior with respect to survival.

Clearly one of the most useful and important aspects of our study of neonatal ECMO has been the lively discussion generated regarding adaptive randomized clinical trials and the need for randomized trials in clinical studies. We certainly have come full circle to the point where a respected statistician suggests that the evidence for superiority of a given treatment can be just as compelling based on historical controls or simply the strength of numbers. We made the same point in a 1988 publication reporting 715 patients treated with neonatal ECMO in a national registry (Toomasian, Snedecor, Cornell, Cilley and Bartlett, 1988). Eighty one percent of these patients survived. This survival rate alone, combined with the number of cases, means that ECMO is statistically significantly better than any other treatment with a survival rate less than a 78.4% (one-sided 95% confidence bound).

In 1982, however, no such argument could be made. We had personal experience with 55 cases but there was very little other experience in the world to support our observation that survival could be improved with ECMO. Of course, we entered this study with the anticipation that the survival rate with ECMO would be much higher than that which had been observed under conventional intensive therapy. We selected a stopping rule that ensured a high probability of selecting ECMO for further study if in fact ECMO were by far the superior treatment. Any clinical study begins with an estimation of the probability of results in both arms of the study. We would call this prior specification of probabilities, for planning purposes, and hypothesis formation rather than bias. We would refer to our study as a well-designed experiment, not a demonstration. It is not necessary to have individual or institutional equipoise to design and conduct a valid clinical experiment.

Despite our anticipation that ECMO would lead a higher survival rate than conventional therapy, we could not be sure that ECMO would be this successful in a clinical trial in which neonates, who met well-defined entry criteria for respiratory distress, were randomized. Evidence of this comes from the fact that two parallel studies were started: one in infants greater than 2 kg and one in infants between 1–2 kg. The latter study was stopped because of intracranial bleeding with ECMO. The anticipated success rate with ECMO was achieved only in the larger neonates. If these studies had

Robert H. Bartlett is Professor of General and Thoracic Surgery at the University of Michigan. His mailing address is 2920 Taubman Health Center, 1500 E. Medical Center Drive, Ann Arbor, Michigan 48109-0331. Richard G. Cornell is Professor of Biostatistics at the University of Michigan. His mailing address is 109 South Observatory Drive, Ann Arbor, Michigan 48109-2029.

been only demonstration trials, as described by Royall, success with ECMO would have been achieved in both studies.

Royall not only fails to realize that the Michigan ECMO trial was an experimental trial, but he also fails to fully describe the evidence in favor of ECMO from that study. As emphasized by Cornell (1989), the Michigan study was a two-part study, with adaptive randomization in the first part of the study and continued study of the treatment selected in the first part (ECMO) during the second part. Cornell, Landenberger and Bartlett (1986) reported 19 successes without any failures with ECMO during both parts of the study. In 1984 the neonatal ECMO Registry was established at the University of Michigan. At that time approximately 200 ECMO cases had been done, most of them by the University of Michigan team. As new ECMO centers became established, they joined the registry. As mentioned earlier, the registry was updated to 715 cases in 1988 with a survival rate of 81%. We mention these results to emphasize that the initial Phase I studies with ECMO, the randomization and continuation parts of the Michigan study, and the registry form a continuum with the study design changing in response to accumulated information for both ethical and scientific reasons.

We acknowledged in our publication that we accepted what Royall calls the practical argument for a randomized clinical trial. There is a cost in doing such a trial, but the cost is exceeded by the benefits to many more patients whose physicians will have been convinced by the randomized trial not to use the inferior treatment.

The reasons that we felt "compelled" to proceed with the randomized trial, using the practical argument, was, as Royall surmises, that none of our colleagues—our own neonatology staff, referring physicians, other prominent neonatologists, other prominent life support researchers, hospital administrators, insurance carriers, the NIH study sections which reviewed our grant applications, the editors which reviewed our publications, and the statisticians which reviewed our claims and methods—believed that ECMO had been tested let alone proven in neonatal respiratory failure. Since our group had first described the technique, and since we had refined the application through utilization in 55 Phase I patients, we ourselves felt compelled to evaluate the new technique in a way that our colleagues listed above acknowledge and respect. We knew that approximately 90% of patients assigned to the control group would die with continuing conventional therapy. This did indeed pose a serious ethical question, which we considered jointly with our NIH study section and our Institutional Review Board. The use of the randomized-play-the-winner statistical method did not eliminate this dilemma, but certainly softened it considerably to the point where the practical argument would prevail.

Excellent statisticians and neonatal clinicians criticized our study because of its adaptive design. Although the survival rate under ECMO was high, the lack of concurrent testing with control patients led them to doubt the findings. However, many physicians did believe the results of the study and began to refer neonatal patients for ECMO. Many major neonatal centers believed the results of the study and set out to establish ECMO treatment teams in their own institutions. Only the Harvard group decided that additional experimentation with randomization to conventional and ECMO therapy was necessary, for reasons described by Ware (1989). Their study was criticized for being unnecessary and unethical, just as ours was criticized as being unbelievable and unethical.

Royall suggests that a new treatment such as neonatal ECMO could have been evaluated by comparing its use in one medical center to results in another medical center, such as Johns Hopkins, where the treatment was not used. Clinicians would be very skeptical that classification and treatment of patients with acute physiologic disorders could be matched closely enough to carry out a believable study. This approach would be reasonable for a condition which is uniformly recognized as fatal such as cancer of the pancreas. A new treatment of biopsy-proven cancer of the pancreas which resulted in 80% survival would be accepted on the basis of an interinstitutional trial because cancer of the pancreas is 95% fatal in every institution where it is treated, and the condition is an anatomic and histologic condition rather than a physiologic condition.

A study similar to that proposed by Royall was done except that controls were studied retrospectively instead of prospectively. Ware (1989) describes an examination at Harvard of records of neonates with severe respiratory distress treated conventionally in 1982 and 1983. Eleven out of 13 died. Despite a comparison of these results with those from the Michigan study, the team at Harvard was not convinced that ECMO had been proven to be more effective than conventional therapy. Thus it is unlikely that the study design suggested by Royall would have led to greater credibility than the Michigan study. It would share with the Michigan study the criticism that the design was inadequate to avoid further randomization to conventional therapy.

Another problem with the study design proposed

by Royall concerns informed consent. Royall objects to the use of the Zelen randomization procedure in both the Michigan and Harvard studies. In his comments on the Harvard study (Royall, 1989), he criticized that study because "parents of a critically ill infant, for whom conventional therapy held little hope of survival, were not even informed that a highly promising alternative therapy was available for their baby." Would the parents of infants treated at Johns Hopkins in his proposed study have been informed about the ECMO procedure being used at Michigan in the other arm of the study? It is likely that consent based on full information on both arms of the study would have led to enough discontent among parents and patient care personnel at Johns Hopkins to have led to discontinuance there. Thus the study mentioned by Royall likely would have been neither credible nor feasible. More basically, a closer look at his suggested study design illustrates the difficulties we faced in the design of our study.

In summary, Royall presents an excellent review of the problem, culminating in the suggestion that our prospective randomized clinical trial was probably not necessary as evidence after all. Some day soon we will undertake studies with the hope of obtaining convincing evidence that ECMO is better than conventional treatment in pediatric respiratory failure and adult respiratory failure. If we propose an approach without randomization to conventional therapy, we certainly will consider calling on Dr. Royall to argue our case with the NIH study section, the insurance carriers and editors of scientific journals.

Comment

David P. Byar, M.D.

1. INTRODUCTION

Voltaire is said to have stated that the price of liberty is eternal vigilance. I often feel that the same applies to defending the use of randomized clinical trials (RCTs) because there are incessant attempts to replace them with other forms of investigation for various reasons. Here the reason has to do with the possibility that many trials may be unethical because of the "personal care principle."

But what is Royall really trying to say? He is clearly not opposed to randomized trials in all circumstances because he urges statisticians to promote randomized trials when they are ethically and practically feasible, and he concludes that "it is important that we understand, teach, and exploit the advantages of RCTs." However, the examples and quotations he chooses and his general emphasis leaves the strong impression that randomized trials are frequently unethical, if not at the outset, then as the data begin to accrue. His stated goal seems to be to encourage statisticians to be more sensitive to ethical issues when they are designing, reviewing or criticizing clinical studies. Few would disagree with this position, but along the way he examines the underlying statistical reasoning that leads to randomization and finds it unconvincing. He also believes that the value of nonrandomized studies has been greatly underestimated and that, because of the ethical dilemma, statisticians should devote more attention to examining alternatives to RCTs. These are very important issues because the use of randomized trials has become a standard and widespread method of evaluating therapies. According to the CLINPROT computerized registry, there are currently some 3,000 randomized trials of cancer treatment. The overall total, including research on other diseases, could easily exceed twice that number.

David P. Byar, M.D., is Chief, Biometry Branch, Division of Cancer Prevention and Control, National Cancer Institute, Executive Plaza North, Suite 344, Bethesda, Maryland 20892.

2. ECMO EXAMPLE IS VERY ATYPICAL

The only trial he discusses in detail is the ECMO study, an extremely atypical trial that has already been the subject of considerable comment. Curiously, Royall fails to point out that a previous issue of this very journal contained an article by Ware (1989) with eight sets of comments by other statisticians including Royall himself who found that study "deeply disturbing." In his two pages of comments on Ware's paper he found it impossible to present all the ethical and statistical arguments