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

that led him to the opinion that that trial was unethical. Presumably the present paper is his effort to rectify that situation, but we need to be careful to assess how far the conclusions he draws in that specific situation apply to other settings.

The atypicality of the ECMO trial is revealed by the title of that exchange: "Investigating Therapies of Potentially Great Benefit: ECMO." Further commentary has been provided recently in an editorial in the *New England Journal of Medicine* (Lantos and Frader, 1990) where the focus is on the ethics of clinical research in pediatrics and the long-term neurological sequelae of ECMO, a subject not addressed by Royall. So, what are the principal ways in which the ECMO trial was atypical? It was designed to evaluate a therapy of very great promise, it was very small (10 patients), it was very unbalanced because an adaptive allocation scheme was used (only one patient got conventional therapy), it was conducted in new-born infants, blinding could not be considered because of the nature of the treatment, and the high risk of early death means that patients who failed on the inferior arm could not benefit from the results of the trial. In addition, Zelen's controversial randomized consent design (Zelen, 1979, 1990) was used.

3. AVOIDING SELECTION BIAS IS TOO LIGHTLY DISMISSED

Although Royall is clearly familiar with the need to avoid selection bias, he dismisses this point rather breezily by noting that statistical adjustment procedures can be used to remove any such biases, even though he admits that this is not an option for unknown prognostic factors or for covariates on which we have no information. I would argue that at best adjustment procedures are only partially effective in removing biases, and that doubts always remain. I have summarized many of the reasons in an article arguing that data bases should never replace RCTs (Byar, 1980).

Anyone who has attended meetings of surgical specialities is likely to hear various leading physicians presenting their personal series of cases and stating that "We get excellent results in carefully selected patients." It is quite appropriate for surgeons to select for operations only those patients whom they feel will benefit, but this is of little help in deciding whom to believe. After all, who is carefully selecting the control patients with which such series are compared? I think Royall has greatly underestimated the problems of analyzing observational studies, but I agree with him that we can sometimes learn without randomization. The question is not whether RCTs provide the *only* valid inferences, but rather what study design should be chosen in particular situations. Green and I have summarized in a hierarchy our beliefs about the strength of evidence concerning treatment efficacy provided by various study designs (Green and Byar, 1984). Least convincing are anecdotal case reports, case series without controls and series with literature controls. Next come analyses using computer data bases, case-control observational studies (Horwitz and Feinstein, 1981a, b) and series with historical controls (Gehan and Freireich, 1981). Most convincing is a RCT, especially if the results are confirmed by more than one trial. Some statisticians would argue that properly conducted meta-analyses are the most convincing of all.

Royall's argument that epidemiologists routinely find it impossible to use randomization but have nevertheless made discoveries of enormous importance is unconvincing because it fails to recognize that exposures in epidemiology are generally not chosen in order to affect some outcome (e.g., before the relationship between cigarettes and lung cancer was established, people did not smoke just to see if they could get lung cancer), but treatments are generally chosen with the specific hope that they will affect relevant outcomes, and thus confounding in observational data is much more likely to be a serious and perhaps insurmountable problem when one is trying to evaluate treatments.

4. WHEN TO RANDOMIZE

Royall divides trials into two categories, demonstration and experimental, and argues that demonstration trials are unethical. I certainly agree that in some situations it would be unethical to conduct a trial, but in specific instances this is often a very difficult decision as it must have been in the case of ECMO. The reason is that most trials do not fit neatly into these two categories. Is this distinction really doing anything more than saying that ethical trials are ethical, but unethical trials are not?

I have recently been working with 21 experienced statisticians on a paper concerning design issues in AIDS trials. We agreed to a set of criteria that should be considered when deciding whether an uncontrolled phase III trial of a new therapy could be justified. Among these were the requirements that sufficient data exist to assure that patients not receiving the new therapy would have a poor prognosis, that side effects sufficient to outweigh the potential benefit not be expected, that the anticipated benefit be sufficiently large that the interpretation would be unambiguous and that the scientific rationale for the treatment be suffi-

ciently strong that a positive result would be widely accepted. These criteria suggest that trials designed mainly to demonstrate to others the benefits of a new therapy already thought to be effective may indeed be unethical. This does not mean that I am opposed to all confirmatory trials. In fact, it is rare that a positive result from a single trial is convincing unless the demonstrated benefit is both very large and very precisely estimated.

Randomization should be used in those situations where there is genuine uncertainty about which of two or more treatments is preferable. This concept, sometimes called the "uncertainty principle," is closely related to the notion of equipoise (Freedman, 1987), but the emphasis is on individual rather than group uncertainty. Trials should be open only to patients for whom there is substantial uncertainty about what treatment to recommend. If the responsible physician or the patient is reasonably certain in the light of current evidence that one of the possible trial treatments is inferior or otherwise inappropriate for that particular patient, then that patient should not be randomized.

In situations where the future treatment of most patients in a trial will depend on the trial results, ethical problems with randomization are considerably diminished because patients in both arms can benefit. For example if the majority of the patients on the inferior arm do not die and if it is not too late to treat them, they may then receive the superior treatment.

5. THE PROBLEM OF ACCUMULATING EVIDENCE

Perhaps the most serious ethical problem discussed in Royall's paper is that moderately strong evidence in favor of one treatment may emerge during the course of the trial. This will upset the equipoise that justified undertaking the trial in the first place, although in any particular trial the result may have occurred by chance alone and could lead to an incorrect and perhaps harmful decision. He rejects on ethical grounds the popular solution of revealing outcome data only to a data monitoring committee and argues that experimental trials evolve into demonstration trials as the data emerge. Unfortunately Royall provides no solution to this problem. He is apparently willing to randomize the first patient, but not the second if the outcome for the first is known. It seems to me that this implies that one cannot conduct RCTs at all. Clinical uncertainty may be more widespread than Royall indicates and need not be so fragile that it could be substantially resolved by data on a few patients. Personally I am prepared to accept the idea of a data monitoring committee as a necessary compromise so that reliable data can be obtained.

6. THE NATURE OF ETHICS

Once in a fit of annoyance about overly restrictive ethical distinctions, I remarked sarcastically that "Ethics is simply what you can get by with." When this quote has been repeated by my colleagues it is usually accompanied by a smile, perhaps because it conjures up to some the image of a cynical and ruthless scientist deceiving and exploiting unsuspecting patients for the sake of research, rather than the image I had intended, that of a beneficent and serious researcher frustrated by a maze of regulations, institutional review committees and insistence on overly detailed informed consent procedures. There is some value in my reply, however, because it serves to remind us that a determination of what is ethical depends strongly on prevailing social norms and thus may vary from one time to another and from place to place. It is therefore not surprising that behavior considered perfectly ethical by some people will be found to be unethical by others. Perhaps that is one reason that many people like to debate ethics—everyone can be an expert because one opinion cannot easily be shown to be better than another.

Nevertheless, there is general agreement, in the United States at least, that certain ethical "principles" should govern medical experimentation on human subjects. These include respect for persons, beneficence (Royall's personal care principle), a favorable risk-benefit ratio for contemplated study treatments as agreed on by an institutional review board and the requirement for voluntary informed consent in most situations.

It is interesting to note that these principles do not necessarily apply to situations other than medical research. For example, young and healthy men and women may be drafted to serve in the military in times of crisis without their permission and with no expectation that they will benefit from the experience. Here the balance between respect for individuals and the needs of the community has clearly been tipped in the other direction.

In medical experiments, however, it is generally agreed that the personal care principle must take precedence over the potential benefit to future patients when conflicts between the two arise. This is the basis on which Royall rejects the "utilitarian" argument that, although the individual may suffer, society as a whole will benefit. Judging by how frequently he uses such phrases as "in the physician's best judgement," "what he believes," "in his personal opinion," etc., Royall appears to believe

that all ethical problems in medical experiments will disappear as long as the physicians treating the patients have a clear conscience and do what they think is in the patients' best interests. Of course this means that two different physicians may make opposite decisions, both of them ethical, but he is untroubled by this so long as the physician maintains his competence and has adequate professional knowledge and skills. But what physician is going to perceive that he is incompetent to decide?

I find Royall's arguments very unpersuasive. Rather than always letting the physicians do what they think is best, it seems to me vastly preferable to encourage them to recognize the limitations of their knowledge and conduct randomized trials. Whenever a trial is proposed it usually means that well-informed physicians differ in their opinions and an honest assessment of the evidence is unconvincing. In such a situation a RCT is much more ethical, in my opinion, than for physicians stubbornly to deny uncertainty and do what they individually think is best, unless one believes that ignorance is ethical.

7. FAILURE TO PRESENT SENSIBLE ALTERNATIVES

Royall provides no alternative methodology that clearly satisfies his demanding ethical standards. The only concrete suggestion I could find in his paper was to use concurrent nonrandomized controls to evaluate a new therapy. He finds no ethical problem here because everyone is exercising his best clinical judgement, but it should be obvious that they cannot all be right about which is the better treatment unless the two or more treatments are equivalent.

Other statisticians have tried to deal with the ethical problem of accumulating evidence by developing various sorts of adaptive allocation procedures, but (with the exception of ECMO) they have been little used in practice. Royall labels this solution as utilitarian and rejects it as failing to satisfy the personal care principle. Although Bailar (1976), Pocock (1979), Simon (1977) and Green et al. (1990) have discussed reasons why adaptive allocation is not widely used, Wei and Byar (1986) have suggested that play the winner rules "might be useful in certain specialized medical situations where ethical problems are paramount and one is reasonably certain that time trends and patient heterogeneity are unimportant." However, I now agree with Royall that adaptive allocation is generally not a very satisfactory solution to the ethical problem.

8. FUNDAMENTAL STATISTICAL PRINCIPLES

As an applied statistician, I do not wish to quarrel with Royall about the fundamentals of statistics and the implications of the ancillarity principle in particular. The proper role of randomization in statistical theory has been a difficult issue for a long time and it will not be resolved by any comments from me. For example, Pearson (1937) commented that "The conception of randomization... is both exceedingly suggestive and often practically useful, but perhaps it should be described as a valuable device rather than a fundamental principle."

I am convinced that the most important reason for randomizing is to avoid bias in the selection of patients in the treatment groups to be compared. However, randomization, especially in small trials, does not guarantee that the groups will be balanced for all known and unknown covariates which might affect the outcome, even when stratification has been used. For this reason the first responsibility of a statistician analyzing trial results is to demonstrate that the treatment groups were comparable. If appreciable imbalances are detected, then before drawing any conclusions, an analysis of some sort should be performed to see how attempts to adjust for the imbalances might affect the overall result. It is not clear to me whether or not Royall would agree with this approach.

After discussing some controversial questions about the foundations of statistics, Cornfield (1976) concludes that "The paradox is that a solid structure of permanent value has, nevertheless emerged, lacking only the firm logical foundation on which it was originally thought to have been built."