Statistical Science
1989, Vol. 4, No. 4, 298–340

Investigating Therapies of Potentially Great Benefit: ECMO

James H. Ware

Abstract. Conventional randomized clinical trials provide a powerful and scientifically rigorous method for comparing medical therapies, but designs using 50-50 randomization (or any other constant assignment probabilities) can raise ethical difficulties when there is strong evidence, either a priori or from study data, that one of the therapies may offer great benefit to individual patients. This article describes such a situation, the evaluation of extracorporeal membrane oxygenation (ECMO) for treatment of persistent pulmonary hypertension of the newborn (PPHN).

For many years, the mortality rate among infants with severe PPHN treated with conventional medical therapy (CMT) was 80% or higher. ECMO treatment of PPHN was introduced in 1977 and, by 1985, several centers had reported survival rates of 80% or better in infants treated with ECMO. Only one randomized trial was reported by the end of 1985. This trial used a randomized play-the-winner design. Eleven patients received ECMO therapy and all survived. Only one patient received CMT and this patient died. In part because of the success of ECMO therapy, this trial provided very little comparative data on the two therapies.

The author and medical colleagues reviewed these data in 1985 and concluded that they did not justify routine use of ECMO without further study. A review of historical experience at two Harvard hospitals showed, however, that 13 infants had developed severe PPHN in 1982 and 1983 and 11 (85%) had died. All received CMT. Thus, there was a strong possibility that a randomized trial would show large differences in survival rates in the ECMO and CMT groups.

To balance ethical and scientific concerns, we designed a clinical trial with a treatment assignment procedure that would begin as conventional randomization but switch to a single therapy study when a prespecified number of deaths was observed in either group. This article describes the study design and results. Nineteen (19) patients were enrolled during the randomized phase; 6 of 10 survived in the CMT group and 9 of 9 in the ECMO group. In the second nonrandomized phase, 20 patients were assigned to ECMO therapy and 19 survived. The survival rates in the two treatment groups were significantly different ($P < .05$ with curtailed sampling) and the profile likelihood for the difference in survival rates between the ECMO and CMT groups gave a one-sided 95% confidence interval with a lower limit of .131. We conclude that ECMO therapy increases survival relative to CMT therapy in patients with severe PPHN.

Key words and phrases: Randomized clinical trials, adaptive randomization, play-the-winner, randomized urn designs.

1. INTRODUCTION

Extracorporeal membrane oxygenation (ECMO) is an external system for oxygenating the blood based on techniques used in cardiopulmonary bypass technology developed for cardiac surgery. As the name suggests, blood is removed from the body via a venous line and passed across a membrane where it is exposed to high concentrations of oxygen. It is then rewarmed and returned to the aorta, thereby functionally bypassing both lungs and heart (Bartlett, 1984; Figure 1). Physicians have been exploring ways to use

James H. Ware is Professor of Biostatistics at the Harvard School of Public Health, 677 Huntington Avenue, Boston, Massachusetts 02115.

ECMO for treatment of acute respiratory conditions for more than two decades. Early efforts to use ECMO for pulmonary support of adult patients with acute respiratory failure led to a clinical trial that failed to demonstrate an improvement in survival (Zapol, Snider, Johnson et al., 1976). In the late 1970's, physicians began to investigate the use of ECMO in the treatment of newborn infants with lung disease. Studies focused on infants with diaphragmatic hernia, an anatomic abnormality that leads to impaired intrauterine growth of at least one lung, and persistent pulmonary hypertension in the newborn (PPHN), a condition characterized by pulmonary hypertension, low blood flow through the lungs, and, as a result, inadequate oxygenation of the blood.

PPHN could be an ideal setting for the application of ECMO therapy. Infants with severe PPHN are at very high risk of death in the first days of life, but many are full-term infants otherwise in good health, so that those who survive have a good long-term prognosis. The therapeutic task is to bring the infant through an acute period of pulmonary insufficiency with minimal neurologic and other consequences resulting either from oxygen deficit or therapy.

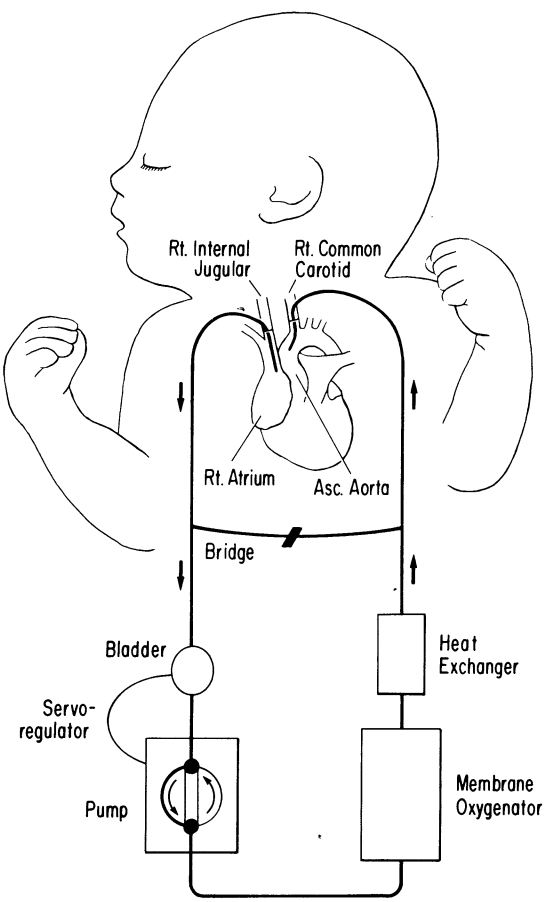

FIG. 1. *Extracorporeal membrane oxygenation, shown schematically, is in essence a simplified heart-lung machine. Cannulas draw blood from the right atrium and return it to the aorta, thereby bypassing both heart and lungs.*

Several centers have reported excellent results in the use of ECMO to treat PPHN (Wetmore, McEwen, O'Connor and Bartlett, 1979; Bartlett, Andrews, Toomasian et al., 1982; Kirkpatrick et al., 1983), but these studies did not include concurrent control groups. Some investigators reported survival rates in excess of 80% in populations of infants believed to have survival rates of 20% or less in the absence of ECMO treatment. The assumption of low survival rates in untreated infants was based on historical experience, and the data supporting these rates were not presented. Nevertheless, medical observers agreed that the infants selected for treatment were at high risk.

Dr. Robert Bartlett and his colleagues at the University of Michigan were among the leaders in exploring this new use of ECMO technology. They recognized that studies involving concurrent control groups would strengthen the evidence for the efficacy of ECMO, but they had strong ethical concerns about studies involving balanced randomization of infants to ECMO and to conventional therapy involving intensive ventilatory support. Specifically, they were concerned about continuing to randomize infants when accumulating evidence strongly suggested the superiority of one treatment. To address these concerns, the investigators designed a clinical trial using adaptive randomization (Bartlett et al., 1985; Cornell, Landenberger and Bartlett, 1986).

The design had several interesting features. First, it used the randomized-consent design proposed by Zelen (1979). Only patients randomized to ECMO therapy were approached for informed consent. The randomized-consent design is attractive in this setting because a standard approach to informed consent would require that parents of infants near death be approached to give informed consent for an invasive surgical procedure that would then, in some instances, not be administered. Those familiar with the agonizing experience of having a child in a neonatal intensive care unit can appreciate that the process of obtaining informed consent would be both frightening and stressful to parents. Second, treatment assignments were based on the randomized urn designs of Wei and Durham (1978). These adaptive designs use randomization probabilities that correspond to sampling from an urn with replacement and additions.

In the study of Bartlett et al., the probabilities corresponded to sampling from an urn which initially contained one ball of each color. Treatments were assigned sequentially by selection *with replacement plus addition* of a ball to the urn after each treatment assignment (Cornell, Landenberger and Bartlett, 1986; Wei, 1988). When a treatment was selected and the infant survived, a ball representing that treatment was added to the urn. When the infant died, a ball representing the other treatment was added. Thus,

the randomization probabilities changed adaptively over time. The total number of patients was fixed in advance and, at the end of the study, the treatment with the better survival rate was selected as the therapy of choice.

The investigators based their sample size calculations on a selection paradigm. By calculating the probabilities associated with different sequences of treatment assignments and outcomes, they showed that, if the survival rates in the two treatment groups differed by at least .40, a study with a total sample size of 10 would have probability at least .95 of selecting the superior therapy. Thus, the trial was designed to enroll 10 infants, with treatment assignments determined by the randomized urn procedure.

This design resulted in a trial in which only one infant received conventional therapy (Bartlett et al., 1985). The first treatment assignment was to ECMO, and the infant survived. The second was to conventional therapy and the infant died. Eight subsequent infants were randomized to ECMO and all survived. Two additional patients were treated with ECMO after the sample size goal had been reached and these patients also survived. The results for all twelve patients were published (Bartlett et al., 1985) along with discussion (Paneth and Wallenstein, 1985; Ware and Epstein, 1985) and a paper from another center reporting 100% survival in 15 patients with PPHN treated with a modified version of conventional ventilatory therapy (Wung, James, Kilchevsky and James, 1985). The discussants expressed concern about the assignment of only one infant to conventional therapy in the study of Bartlett et al. and noted that this patient was the most severely ill patient in the study.

In retrospect, the Michigan study provided encouraging information about survival rates of infants treated with ECMO, but gave very little information about survival rates in the same population treated with conventional therapy. The investigators reported that historical experience in their center would suggest a survival rate of 20% or less in infants meeting study eligibility criteria, but did not provide data on historical controls.

After reviewing these results and available data from ECMO registries, the author and medical colleagues at Boston's Children's Hospital Medical Center (CHMC) and Brigham and Women's Hospital (BWH) concluded that the data were not sufficient to justify routine use of ECMO in the treatment of PPHN. We were uneasy about rapid acceptance of a new and potentially dangerous technology based on inadequate experience from randomized clinical trials, but shared the concerns of the Michigan group about the ethical difficulties that might arise in a new randomized trial if early experience again suggested that ECMO therapy was dramatically effective in this group of very high risk infants. This period was one of intense debate between proponents of ECMO, who believed that the therapy was a breakthrough in treatment of PPHN, and skeptics, who were unconvinced by the registry data on mortality rates and expressed concerns about potential morbidity of ECMO treatment, especially brain hemorrhage and subsequent severe impairment.

Ethical issues about randomization arise in many clinical trials. Freedman (1987) defines *equipoise* as a state of genuine uncertainty about which of two therapies is superior. He notes that a state of equipoise can be disturbed by accumulating data in a clinical trial, raising ethical issues about further randomization. Ethical concerns about randomization can be greatly intensified when the difference in efficacy of the two therapies may be very great. In that instance, the outcome for the next study patient may depend primarily on the treatment assignment.

While exploring design strategies that would balance ethical and scientific concerns, the author discussed designs based on a maximum number of deaths in either treatment group with Dr. Marvin Zelen. Work stimulated by that discussion led to the design and clinical trial described in this paper.

2. STUDY DESIGN

Consider a family of study designs characterized by constraints on the maximum number of deaths, r, allowed in either treatment group. (The theory is easily extended to allow different maximum values in different treatment groups.) Initially, patients are assigned to treatments by any randomization procedure. In the study discussed here, treatments were selected by a randomized permuted blocks design with blocks of size four. When r deaths occur in one of the treatment groups, randomization ceases and all subsequent patients are assigned to the other treatment until r deaths occur in that arm or until the number of survivors is sufficient to establish the superiority of that treatment arm, using testing procedures described below.

Let s_1 and s_2 be the total numbers of survivors in the two treatment arms during the randomized and nonrandomized phases. Then s_i is distributed as a negative binomial random variable with probability p_i, the probability of survival for patients receiving treatment i. If $s = s_1 + s_2$, the conditional distribution of s_i given s is

$$P(s_1, s_2 \mid s) = \frac{\binom{r+s_1-1}{s_1}\binom{r+s_2-1}{s_2} p_1^{s_1}(1-p_1)^r p_2^{s_2}(1-p_2)^r}{\sum_{u=0}^{s}\binom{r+u-1}{u}\binom{r+s-u-1}{s-u} p_1^{u}(1-p_1)^r p_2^{s-u}(1-p_2)^r}. \tag{1}$$

When $p_1 = p_2$, this distribution is parameter-free. Otherwise, it depends on the survival probabilities only through the relative risk, p_1/p_2. Thus, tests for the equality of survival rates and confidence intervals for the relative risk can be based on this conditional distribution. If the study is stopped early, however, the statistics required for the confidence interval calculation will not be observed.

The power of the study for detecting different survival rates depends on r and on the hypothesized survival rates in the two treatment groups. A chart review identified 39 patients with PPHN treated at CHMC or BWH in 1982 and 1983. A total of 11 (85%) died in a subgroup of 13 patients with severe PPHN (persistently low aortic oxygen concentrations during the interval between 12 and 72 hours after birth while receiving maximal medical therapy). These criteria, consistent with accepted definitions of severe disease, were used as eligibility criteria for the randomized trial. A subsequent comparison with two other ECMO studies showed that the criteria used in the three studies were virtually equivalent in determining eligibility of individual patients (O'Rourke et al., 1989).

Based on this review and data from previous studies, sample size calculations were based on the following null and alternative hypotheses:

$$H_0: p_1 = .20, \quad p_2 = .20$$
$$H_1: p_1 = .20, \quad p_2 = .80$$

where p_1 and p_2 are the survival rates in the CMT and ECMO groups, respectively. Unconditional power calculations based on (1) and the marginal distribution of s, the total number of survivors, showed that a design based on a maximum of four deaths in each treatment group and a Type I error rate of .05 (one-sided) would have power of .77 against H_1 if a conditional test based on (1) was used. Under the same conditions, the design based on three failures would have power of .55 and five failures would have power of .84. The study was therefore designed to discontinue randomization when the fourth death occurred in either treatment arm.

3. RESULTS

The trial began on February 6, 1986. Patients were randomized in blocks of four, and treatments were assigned randomly to the first 19 patients. Of these 19 patients, 10 received CMT, including patient 19, and 4 died. The remaining 9 patients received ECMO and all survived.

Randomization ceased at this point, but accrual to ECMO therapy continued. Investigators responsible for recruitment and patient care were not immediately told that randomization had been discontinued. Because treatment could not be blinded, however, it soon became apparent that study patients were consistently receiving ECMO. The investigators then met to discuss the need to continue enrolling patients in the study, as well as the importance of continuing recruitment and patient management procedures used during the randomized phase of the study. One eligibility criterion was inability to treat the PPHN effectively using maximal ventilatory support. This criterion prevented the admission of less severely ill patients in the second phase of the study.

Given six survivors in the CMT group, the conditional distribution (1) implied that the survival rates in the two treatment groups would be significantly different if at least 28 patients survived ECMO therapy before the fourth death in that treatment arm. Thus, the study was continued with the goal of assigning infants to ECMO until either the 28th survivor or the 4th death was observed. Twenty additional patients were enrolled over a period ending on July 1, 1988, and 19 of these patients survived. The trial was terminated at that time.

In total, 39 patients were enrolled. The survival experience of these patients is shown in Table 1. Table 2 gives selected patient characteristics of CMT and ECMO patients during the randomized phase and of ECMO patients during the nonrandomized phase of the study. A more extensive comparison is given elsewhere (O'Rourke et al., 1989). Only one statistically significant difference in the distributions of patient characteristics in the three groups could be identified. Patients enrolled in phase 2, the nonrandomized phase, had a significantly higher age at enrollment than those enrolled in phase 1, presumably because more of these patients were outborn. The data suggest that outborn patients in the CMT group were at higher risk of death.

During the period of study, five infants hospitalized at either CHMC or BWH met eligibility criteria for the study but were not enrolled. In three cases, physician or technician coverage for ECMO therapy was not available. In two other cases, all ECMO machines were in use. Two of these patients died and two survived after conventional therapy; one patient received ECMO at another hospital and survived.

Because sampling was curtailed before the fourth death in the ECMO group, only an upper bound ($P < .05$) could be calculated for the significance level

TABLE 1

Survival experience of patients randomized to ECMO and conventional therapy (CMT) during phase 1, the randomized phase of the trial, and phase 2, the nonrandomized phase

	Phase 1		Phase 2	
	ECMO	CMT	ECMO	CMT
Lived	9	6	19	0
Died	0	4	1	0

TABLE 2
Characteristics of patients randomized to ECMO and conventional therapy (CMT) during phase 1 and of patients assigned to ECMO during phase 2

Variable	Treatment Group		
	CMT phase 1 ($n = 10$)	ECMO phase 1 ($n = 9$)	ECMO phase 2 ($n = 20$)
Birth Weight	3.40 (.09)[a]	3.62 (.13)	3.38 (.13)
Gestational Age	39.4 (.5)	40.4 (.4)	39.8 (.3)
Apgar 1	5.4 (.8)	5.9 (.7)	4.8 (.7)
Apgar 5	7.4 (.5)	7.3 (.7)	6.8 (.6)
Age at entry (hrs)	26.2 (4.5)	20.0 (2.0)	33.9 (3.5)
Diagnosis [deaths]			
Meconium aspiration	4 [2]	8	11 [1]
Sepsis	1	1	3
Idiopathic PPHN	5 [2]	0	6
Place of birth [deaths]			
Outborn	5 [4]	4	15 [1]
Inborn	5	5	5

[a] Mean (standard error).

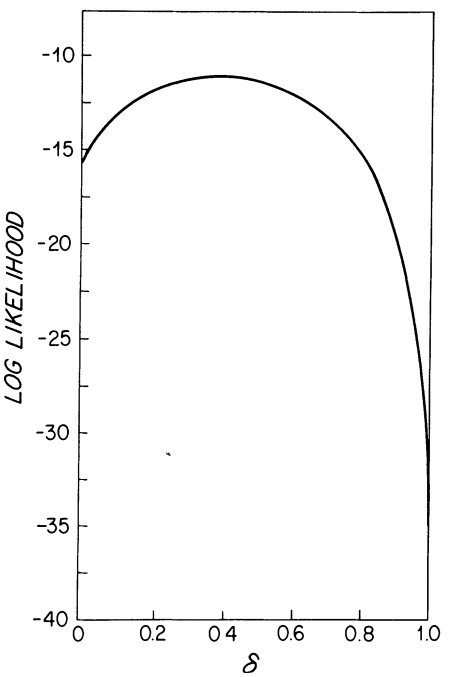

FIG. 2. *Profile likelihood for $\delta = p_2 - p_1$, the difference between the survival rates in the ECMO and CMT groups.*

that would have been achieved if the study had been continued to the fourth death. Moreover, distribution (1) could not be used to calculate a confidence interval for the risk ratio. The validity of confidence intervals based on profile likelihoods is not, however, affected by curtailment of sampling. Thus, we calculated the profile likelihood for the risk difference, $p_2 - p_1$, the parameter of medical interest (Wei, Smythe and Park, 1988, Figure 2). Large sample theory for the likelihood ratio gives a one-sided 95% confidence interval with a lower limit of .131 for $p_2 - p_1$.

4. DISCUSSION

This discussion addresses four issues, 1) the appropriateness of a study with a nonrandomized phase in this setting, 2) the adequacy of the sample size, 3) the potential utility of other designs such as adaptive or sequential methods, and 4) the implications of the study for future use of ECMO.

Some colleagues have argued that randomization should have continued even after four deaths had occurred in the conventional therapy arm. This position is easy to defend on scientific grounds. Designs using randomization blocked on potentially important confounding variables, including time of enrollment, remain the "gold standard" for comparison of therapies (Feinstein, 1988). Patients enrolled in the second phase of the ECMO study did not have concurrent controls, raising the possibility of noncomparability of treatment groups. The three patient groups were comparable with respect to measured characteristics (Table 2), but such comparisons do not guarantee the comparability of the treatment groups. The issue is similar to that faced in the interpretation of studies using historical controls. In this instance, of course, the second phase directly followed the first and all enrollment, treatment, and data collection procedures were on a common protocol.

The arguments in favor of the two-phase design are principally ethical and touch on an issue that arises in any clinical trial which ultimately shows the superiority of one therapy. The ethical issues are intensified, however, when the difference in efficacy of the two therapies is large.

First, consider the data provided by the 19 randomized patients. Fisher's exact test gives a P value of .054 for the comparison of 4 deaths in 10 patients to

0 deaths in 9 patients. Thus, the difference in survival rates was nearly significant when randomization ceased. We return to this point subsequently.

A Bayesian analysis of these 19 cases provides another way of looking at the ethical dilemma. Consider a prior probability distribution that assigns ⅓ of the prior probability to each of three regions: $p_1 < p_2$, $p_1 = p_2$, and $p_1 > p_2$. A prior of this type is given below. Such a prior can be combined with the accumulated data to calculate the posterior probability that ECMO is inferior to CMT.

We assume that p_1 and p_2 have a joint prior probability distribution of the following form:

1) p_1 has a beta distribution with parameters a and b.
2) The conditional distribution of p_2 given p_1 is given by

$$P(p_1 < p_2) = P(p_1 = p_2) = P(p_1 > p_2) = \tfrac{1}{3},$$

and

$$f(p_2 \mid p_1 = p_1^*, p_2 < p_1) = (p_1^*)^{-1},$$
$$f(p_2 \mid p_1 = p_1^*, p_2 > p_1) = (1 - p_1^*)^{-1}.$$

This family of prior distributions assigns nonzero prior probability to the null hypothesis, as suggested by Jeffreys (1948) and Cornfield (1966), and assigns the remaining probability equally to regions of superiority of ECMO and conventional therapy. It is conservative, in that the investigator's prior gave greater weight to the superiority of ECMO than of CMT.

If

$$P(p_1) = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} p_1^{a-1}(1-p_1)^{b-1}$$

then the prior probability for any region, R, in the unit square is the sum of the two-dimensional integral of the function

$$P(p_1, p_2) = \begin{cases} \frac{1}{3}\frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} p_1^{a-2}(1-p_1)^{b-1}, & p_1 > p_2, \\ \frac{1}{3}\frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} p_1^{a-1}(1-p_1)^{b-2}, & p_1 < p_2 \end{cases}$$

over all of R except the 45 degree line and the one-dimensional integral of the function

$$P(p_1) = \frac{1}{3}\frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} p_1^{a-1}(1-p_1)^{b-1}$$

along the 45 degree line within R.

The likelihood function based on the first 19 observations is $p_1^6(1-p_1)^4 p_2^9$. The posterior probabilities are

$$P(p_1 > p_2) = \frac{F_1}{F_1 + F_2 + F_3},$$
$$P(p_1 = p_2) = \frac{F_2}{F_1 + F_2 + F_3},$$
$$P(p_1 < p_2) = \frac{F_3}{F_1 + F_2 + F_3},$$

where

$$F_1 = \int_0^1 \int_0^{p_1} p_1^{a-2}(1-p_1)^{b-1} p_1^6 (1-p_1)^4 p_2^9 \, dp_1 \, dp_2$$
$$= \int_0^1 p_1^{a+4}(1-p_1)^{b+3} \int_0^{p_1} p_2^9 \, dp_1 \, dp_2$$
$$= \frac{1}{10}\int_0^1 p_1^{a+14}(1-p_1)^{b+3} \, dp_1$$
$$= \frac{1}{10}\frac{\Gamma(a+15)\Gamma(b+4)}{\Gamma(a+b+19)}$$

and, similarly

$$F_2 = \frac{\Gamma(a+15)\Gamma(b+4)}{\Gamma(a+b+19)},$$
$$F_3 = \frac{1}{10}\left[\frac{\Gamma(a+6)\Gamma(b+3)}{\Gamma(a+b+9)} - \frac{\Gamma(a+16)\Gamma(b+3)}{\Gamma(a+b+19)}\right].$$

If the prior for p_1 is uniform on [0, 1], then $a = b = 1$, and the posterior probabilities are $P(p_1 > p_2) = .01$, $P(p_1 = p_2) = .10$, and $P(p_1 < p_2) = .89$. Given the historical experience in 13 patients, the beta prior with $a = 3$, $b = 12$, is also of interest. This prior is consistent with reports from other centers of 80% mortality in conventionally treated patients. With this prior, the posterior probabilities are $P(p_1 > p_2) = .0004$, $P(p_1 = p_2) = .0039$, and $P(p_1 < p_2) = .9957$. Thus, the posterior probability that ECMO is inferior to conventional therapy, given experience in 19 randomized patients, is only .01 when one chooses a prior that ignores recent experience with conventional therapy and falls to .0004 when one bases the prior on that experience.

The experience in the 13 selected patients is affected by selection bias, because the data were explored for subgroups with high mortality. Nevertheless, the posterior probability that ECMO has a higher mortality rate than CMT is less than .01 and the posterior probability that ECMO has a lower mortality rate is at least .90 over a wide range of priors. Given this analysis, it is difficult to defend further randomization ethically.

In retrospect, the assumptions used in the sample size calculations seem somewhat optimistic. The

possibility that the survival rate of patients treated with CMT was underestimated from the historical data has already been noted, though the survival rates in the historically selected group of 13 and the 10 patients treated with CMT during the study are not significantly different. We used a one-sided testing procedure and power of .77 to choose a relatively small value for r, the maximum number of deaths. This small value appealed to the concern that ECMO might prove to be dramatically more effective than conventional therapy, a concern that seems appropriate in view of study data. The sample sizes required to test the specified null and alternative hypotheses at these error rates were also achievable within the context of a single-institution study.

Ironically, a conventional randomized clinical trial testing the same null and alternative hypotheses with Type I and Type II error rates of .05 and .20, respectively, would have required only 10 patients in each group (Haseman, 1978). The first 19 patients came close to completing this design. Because of blocking, the 20th patient would have been randomized to ECMO, the therapy actually received. The P value for Fisher's exact test applied to the 2×2 table given by the first 20 patients is .043. We chose not to use this design, however, because of the possibility that eight or more of the patients receiving CMT would die.

The design used in this study is a simple alternative to more sophisticated adaptive designs (Simon, 1977; Bather, 1985). Any adaptive scheme that allows randomization probabilities to depend on accumulating data can, however, introduce bias in comparisons between treatment groups if the patient population changes over time. Moreover, the design used in this study was superior to adaptive designs inducing gradual changes in assignment probabilities in one respect. It avoided the ethically difficult situation of a protracted study period during which randomization probabilities were very unequal, leading to the infrequent assignment of a patient to what appeared to be an inferior therapy. Some statisticians believe that randomization with constant randomization probabilities should be continued so long as randomization is ethically justified, and that adaptive schemes are an insufficient response to evidence that the therapies are not equally effective.

Among the methods for adaptive randomization, the randomized play-the-winner rules of Wei and Durham offer an intuitively appealing approach. This method makes it difficult to compute significance levels and confidence intervals, but Wei (1988) has described efficient algorithms for implementing exact two-sample tests and Wei, Smythe, Lin and Park (1990) have shown that profile likelihood methods give confidence intervals that perform well in moderate samples. Wei calculated an exact one-sided P value of .051 for the study of Bartlett et al. (1985).

Sequential methods are not especially effective in this setting. Although they provide a mechanism for early termination, they do not use adaptive randomization. The Sequential Probability Ratio Test, for example, would terminate only if the first five discordant pairs or eight of the first nine pairs favored ECMO. An Armitage closed sequential plan with a maximum of 13 discordant pairs would stop if the first 6 pairs or 10 of the first 11 pairs favored ECMO (Wetherill and Glazebrook, 1986).

As noted previously, the ethical issues discussed here can arise in any clinical trial in which patient outcomes are observed acutely, so that patient enrollment continues until the study ends. Many clinical trials investigate therapies that have very small differential effects on patient outcome, or focus on endpoints other than mortality. In those situations, it is sometimes argued that the benefits associated with participation in a clinical trial more than offset the small losses associated with receiving the inferior therapy. This argument does not seem relevant here.

Some statisticians argue that, so long as the accumulated data do not demonstrate the superiority of one therapy by the criterion of statistical significance, perhaps adjusted for sequential analysis, the therapies have not been shown to differ in their efficacy, so that there is no reason to discontinue randomization. This argument also seems unsatisfactory in situations where patients may benefit substantially from the better therapy.

The organizational structure for many randomized clinical trials includes a Policy Advisory Board with duties including responsibility for monitoring accumulating data to determine whether the trial should be terminated for ethical reasons. This system has many desirable features, including the separation of responsibility for scientific investigation and patient care. Freedman (1987) argues that *clinical* equipoise, a lack of consensus within the expert medical community about the comparative merits of two therapies, can justify continuation of a clinical trial even when those directly involved in patient care are no longer neutral about the choice between therapies. From this perspective, the Policy Advisory Board can represent the larger scientific community and support continuation of a clinical trial when the investigators are no longer in a state of equipoise. Although we initially felt that a Policy Advisory Board would not be necessary in a small single-institution study such as this one, an external group might have provided valuable input on the difficult ethical and scientific decisions involved in the design and conduct of the study.

Such institutional arrangements can not, however, fully address the ethical concerns that arise when previous studies and accumulating evidence strongly suggest differential effectiveness of study therapies. For studies in which this issue is likely to arise,

adaptive schemes like that described here provide a structured approach to comparing therapies that has many of the strengths of randomized clinical trials.

This study has raised many different statistical issues. First, it represents one of a very few studies that have used randomized consent. We believe that randomized consent was ethically justified in this setting, and the study design was approved by two Institutional Review Boards. Nevertheless, the need to withhold information about the study from parents of infants receiving CMT raises some difficult questions. Second, our design used only the number of deaths in each group. The unexpectedly high survival rate in the CMT group led to a larger study than was initially anticipated. Other adaptive randomization strategies or methods of analysis might offer advantages over the one we chose. The decision to combine a randomized and nonrandomized phase deserves further discussion. We recognize the limitations of this approach relative to the classical randomized clinical trial and made special efforts to maintain other strengths of randomized trials, especially standardized accrual, treatment, and data collection methods, throughout the study. Finally, adaptive schemes such as the one used in this study may have a wider role in the evaluation of new therapies of potentially great benefit.

Although the survival data from this study are compelling, their implications for the use of ECMO in clinical care have important limitations. First, the eligibility criteria for the study were very narrow, and the implications of the study for less severely ill patients and patients with conditions other than PPHN remain to be determined. Second, questions remain about the morbidity of ECMO, especially brain hemorrhage (Shumacher et al., 1988). There was no evidence of differential morbidity during the acute phase of this study, however, and all surviving patients have been enrolled in a comprehensive follow-up program. Data comparing longer-term outcome in survivors treated with ECMO and CMT will be reported as they become available. These data will be limited, of course, by the small number of survivors in the CMT group.

Despite these limitations and the better than expected survival rates of infants treated with CMT, demand for ECMO has grown rapidly at the participating hospitals since the data were reported to the medical staff. As with any costly, technology-intensive therapy, increasing demand for ECMO will raise difficult issues about constraints on medical resources and the management of demand that exceeds the capacity of the medical system.

ACKNOWLEDGMENTS

This work has benefited substantially from conversations with Dr. Marvin Zelen and two medical colleagues, Dr. Michael Epstein and Dr. Pearl O'Rourke. Ms. Catherine Weeks performed the sample size calculations and Ms. Kimberly Gauvreau calculated the profile likelihood. This research was supported in part by National Institutes of Health Grant GM-29745.

REFERENCES

BARTLETT, R. H. (1984). Extracorporeal oxygenation in neonates. *Hospital Practice* April 139–151.

BARTLETT, R. H., ANDREWS, A. F., TOOMASIAN, J. M. et al. (1982). Extracorporeal membrane oxygenation for newborn respiratory failure: 45 cases. *Surgery* **92** 425–433.

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** 479–487.

BATHER, J. A. (1985). On the allocation of patients in sequential medical trials. *Internat. Statist. Rev.* **53** 1–14.

CORNELL, R. G., LANDENBERGER, B. D. and BARTLETT, R. H. (1986). Randomized play-the-winner clinical trials. *Comm. Statist. A—Theory Methods* **15** 159–178.

CORNFIELD, J. (1966). A Bayesian test of some classical hypotheses—with applications to sequential clinical trials. *J. Amer. Statist. Assoc.* **61** 577–594.

FEINSTEIN, A. R. (1988). Scientific standards in epidemiologic studies of the menace of daily life. *Science* **242** 1257–1263.

FREEDMAN, B. (1987). Equipoise and the ethics of clinical research. *New England J. Med.* **317** 141–145.

HASEMAN, J. K. (1978). Exact sample sizes for use with the Fisher-Irwin test for 2 × 2 tables. *Biometrics* **34** 106–109.

JEFFREYS, H. (1948). *Theory of Probability*, 2nd ed. Clarendon Press, Oxford.

KIRKPATRICK, B. V., KRUMMEL, T. M., MUELLER, D. G., ORMAZABAL, M. A., GRENFIELD, L. J. and SALZBERG, A. M. (1983). Use of extracorporeal membrane oxygenation for respiratory failure in term infants. *Pediatrics* **72** 872–876.

O'ROURKE, P. P., CRONE, R. K., VACANTI, J. P., WARE, J. H., LILLIHEI, C. W., PARAD, R. B. and EPSTEIN, M. F. (1989). Extracorporeal membrane oxygenation and conventional medical therapy in neonates with persistent pulmonary hypertension of the newborn: A prospective randomized study. *Pediatrics* **84** 957–963.

PANETH, N. and WALLENSTEIN, S. (1985). Extracorporeal membrane oxygenation and the play-the-winner rule. *Pediatrics* **76** 622–623.

SHUMACHER, R. E., BARKS, J. D. E., JOHNSTON, M. V., DUNN, S. M., SCHER, M. S., ROLOFT, M. D. and BARTLETT, M. D. (1988). Right-sided brain lesions in infants following extracorporeal membrane oxygenation. *Pediatrics* **82** 155–161.

SIMON, R. (1977). Adaptive treatment assignment methods and clinical trials. *Biometrics* **33** 743–749.

WARE, J. H. and EPSTEIN, M. D. (1985). Comments on "Extracorporeal circulation in neonatal respiratory failure: A prospective randomized study," by R. H. Bartlett et al. *Pediatrics* **76** 849–851.

WEI, L. J. (1988). Exact two-sample permutation tests based on the randomized play-the-winner rule. *Biometrika* **75** 603–606.

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

WEI, L. J., SMYTHE, R. T., LIN, D. Y. and PARK, T. S. (1990). Statistical inference with data-dependent treatment allocation rules. *J. Amer. Statist. Assoc.* **85**. To appear.

WETHERILL, G. B. and GLAZEBROOK, K. D. (1986). *Sequential Methods in Statistics*, 3rd ed. Chapman and Hall, London.

WETMORE, N., MCEWEN, D., O'CONNOR, M. and BARTLETT, R. H. (1979). Defining indications for artificial organ support in

respiratory failure. *Trans. Amer. Soc. Artificial Internal Organs* **25** 459–461.

WUNG, J.-T., JAMES, L. J., KILCHEVSKY, E. and JAMES, E. (1985). Management of severe respiratory failure with persistence of the fetal circulation, without hyperventilation. *Pediatrics* **76** 488–494.

ZAPOL, W. M., SNIDER, M. T., JOHNSON, F. W. et al. (1976). Extracorporeal membrane oxygenation in severe acute respiratory failure: A randomized prospective study. *J. Amer. Med. Assoc.* **242** 2193–2196.

ZELEN, M. (1979). A new design for randomized clinical trials. *New England J. Med.* **300** 1242–1245.

Comment: Ethics and ECMO

Donald A. Berry

I will address several general issues that the Ware paper raises. These include the use of historical controls, the ethics of randomized trials, the impracticality of Neyman-Pearson inference, and optimal adaptive design. I will also suggest a more ethical and perhaps more scientific approach to medical research than that of randomized clinical trials.

RANDOMIZED CLINICAL TRIALS: THE EMPEROR'S NEW CLOTHES

Randomization has achieved hallowed status in biostatistics. Some biostatisticians and clinicians refuse to believe that a treatment has an effect unless it has been shown in a "properly conducted" randomized clinical trial. A report of a randomized clinical trial takes for granted that the trial provides the conclusive answer: if its conclusion is the same as the prevailing wisdom that is based on historical data, the authors tell us that we can finally believe this wisdom; if it differs, they chide historical data and extol the virtues of randomized studies. In the case of ECMO, there was a substantial amount of historical data that, in my view, not only carry more weight than the Ware study, but suggest that randomizing patients to non-ECMO therapy as in the Ware study was unethical.

Ware refers to several previous studies concerning ECMO. The Bartlett et al. (1985) study included 12 patients in its play-the-winner phase; all 11 ECMO patients survived and the conventional therapy patient died. Bartlett et al. also reported on 10 patients who met their entry criteria but were treated after the study was completed: all 8 patients assigned to ECMO survived and the 2 assigned to conventional therapy died (though the authors do not indicate the reasons for different therapy assignments—one possibility unrelated to prognosis is the availability of ECMO machines). Bartlett et al. say they admitted only patients who had at least an 80% chance of dying on conventional therapy. I am currently examining historical controls provided by Dr. Bartlett to verify this mortality rate, and so far I have no reason to doubt it. The 40% (4 of 10) death rate on CMT in the Ware study is somewhat inconsistent with an 80% mortality rate, but patients in the Bartlett et al. study generally had worse prognoses than those in the Ware study.

Commenting on the Bartlett et al. study, Ware and Epstein (1985) lament its 50% false-positive rate (or type I error level) since "in trials comparing equally effective innovative and standard therapies, the innovation would be identified as superior therapy in 50% of the trials." Type I error levels do not depend on the data; they are unconditional measures of inference. In particular, they average over data that might have occurred but did not. So the significance level of ½ would apply even if it happened that equal numbers had been assigned to the two therapies with all failures on one therapy and all successes on the other (this is unlikely but possible when using randomized play-the-winner assignment). I will return to conditional versus unconditional inference below. Ware and Epstein conclude 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." Hence the current study.

I disagree with the conclusion of Ware and Epstein: there was ample evidence in the Bartlett et al. study and in other evidence available at the time to conclude that ECMO is beneficial. (And I felt as strongly about this before I became aware of the Ware study.) This is clear if one uses measures of inference that condition on the observed data. For example, a Bayesian analysis that takes into account historical controls and the differing prognoses of the patients shows a dramatic benefit for ECMO (Berry and Hardwick, manuscript in preparation). Historical controls are

Donald A. Berry is Professor, School of Statistics, University of Minnesota, Minneapolis, Minnesota 55455.