[Investigating Therapies of Potentially Great Benefit: ECMO]: Comment

D. Y. Lin; L. J. Wei

Statistical Science, Vol. 4, No. 4. (Nov., 1989), pp. 324-325.

Stable URL:
http://links.jstor.org/sici?sici=0883-4237%28198911%294%3A4%3C324%3A%5BTOPGB%3E2.0.CO%3B2-S

Statistical Science is currently published by Institute of Mathematical Statistics.

Comment

D. Y. Lin and L. J. Wei

As always, it is our great pleasure to read Professor Ware's paper. We appreciate the opportunity to discuss this interesting and well-written article. As far as we know, the Boston ECMO study was the second well-documented trial which used an adaptive design, the first one being the Michigan ECMO study. We regard this work of Professor Ware and his medical colleagues as a very important contribution to the area of clinical trials.

Although various data-dependent treatment assignment rules have been studied for the past three decades, their applications to medical trials are rather limited. As discussed in Professor Ware's paper, most adaptive designs are deterministic and may introduce bias into the study. In the Boston ECMO trial, after the randomization ceased, it soon became evident to the medical investigators that patients were all receiving ECMO. This led Professor Ware to suspect that sicker babies were enrolled in the second phase of the study. Indeed, the patient group in this stage seemed different from that of the first stage. This kind of selection bias reduces the degree of significance of the ECMO effect. Another reason that investigators do not use adaptive designs is that the responses from the study patients are usually not instantaneous. Most data-dependent rules have very little value for such situations.

An interesting adaptive design, the randomized play-the-winner rule of Wei and Durham (1978), was used in the Michigan ECMO study. This rule is a randomized version of Zelen's play-the-winner rule and is applicable to the case where patients may have delayed responses. It is not deterministic and is less vulnerable to experimental bias than other adaptive designs. This design can be easily described with an urn model. An urn has balls of two different types which are marked A and B. We start with α balls of each type. When a patient enters the study, a ball is drawn at random and replaced. If it is type k, then treatment k is assigned to this patient, where $k = A, B$. When a response of a previous patient to treatment k occurs, we add β balls of type k to the urn if the response is a success. Otherwise, β balls of the other type are added to the urn. Let this rule be denoted by RPW(α, β). It can be shown that such designs tend to assign more patients to the better treatment.

The RPW(1, 1) design was used for the Michigan ECMO trial. Partly by chance and partly due to the early successes of the ECMO babies, only one baby was assigned to the conventional treatment. Hence, the Michigan study did not provide sufficient concurrent experience with both regimens. Bartlett et al. (1985) concluded in retrospect that it might have been better to use a RPW(3, 1) design. A RPW(α, β) with $\alpha \geq 3$ gives high probability of having several early subjects on both treatment arms. Similarly, the permuted block design employed in the first phase of the Boston study ensured that there would be sufficient numbers of patients in both groups.

After conducting clinical trials with adaptive designs, it is important to know how to analyze this kind of data. For a Bayesian or other believer in the likelihood principle, the assignment and the stopping rules used for the trial will be ignored in the analysis. However, it is less clear how to analyze these data from a frequentist point of view. Wei, Smythe, Lin and Park (1990) studied a number of inference procedures under the RPW(α, β) design. Their exact methods can be modified to accommodate other adaptive designs, even those including early termination such as the one used in the Boston study. Furthermore, their asymptotic procedures can be applied to other adaptive designs with mild constraints.

Dr. Ware reported a profile likelihood confidence interval for the difference Δ between the survival rates of ECMO and CMT. Such an approximate interval is obtained by inverting the likelihood ratio statistic for testing $\Delta = \Delta_0$. Therefore, a profile likelihood confidence interval is genuinely *two-sided*. The 95% one-sided interval (0.131, 1) reported by Professor Ware was actually based on the 90% two-sided interval (0.131, 0.626). Other two-sided profile intervals with different confidence coefficients are shown in our Table 1.

Wei, Smythe, Lin and Park (1990) show that the profile likelihood intervals perform well for a RPW design with sample size more than 50. However, a design with both adaptive and sequential features was used in the Boston trial, and it is not clear to us if these large-sample procedures are appropriate or not. To this end, a small simulation study was conducted with the present design. Since Dr. Ware's design is an

D. Y. Lin is a Postdoctoral Research Fellow, Department of Biostatistics, Harvard School of Public Health, 677 Huntington Avenue, Boston, Massachusetts 02115. L. J. Wei is Professor, Department of Statistics, University of Wisconsin, 1210 West Dayton Street, Madison, Wisconsin 53706.

TABLE 1
Confidence intervals for the Boston ECMO data

Interval procedure	$1-\gamma$		
	0.99	0.95	0.90
Two-sided profile likelihood intervals for Δ	(0.023, 0.753)	(0.094, 0.672)	(0.132, 0.626)
One-sided exact unconditional intervals for Δ taking account of the design	(−0.180, 1)	(0.035, 1)	(0.148, 1)
One-sided exact conditional intervals for the odds ratio ignoring the design	(0.966, ∞)	(1.834, ∞)	(2.569, ∞)

TABLE 2
Empirical probabilities that the lower bounds of the two-sided profile likelihood confidence intervals for Δ are less than the true Δ under the design of the Boston study

p_A	p_B	Two-sided nominal level		
		0.99	0.95	0.90
0.9	0.1	1.000	1.000	1.000
0.8	0.2	1.000	0.894	0.894
0.7	0.3	0.972	0.972	0.913
0.6	0.4	0.992	0.931	0.918

open one, it does not seem feasible to calculate the empirical coverage probability of a *two-sided* profile likelihood interval for Δ. Consequently, for each simulated sample, we check if the true Δ is larger than the lower bound of the two-sided interval. Each entry in our Table 2 is based on 100,000 replications. Naturally, the reported empirical probabilities that the lower bounds are less than the true Δ's are generally higher than the actual coverage probabilities of the two-sided intervals. The results of this simulation study suggest that the profile likelihood intervals may not be adequate for the present design.

An efficient algorithm has been developed by Wei, Smythe, Lin and Park (1990) to construct exact confidence intervals for Δ under the randomized play-the-winner rule. That algorithm can be easily adapted to the Boston trial. Here, we show how to obtain unconditional confidence intervals for Δ. Let n_A, n_B, s_A and s_B be the observed numbers of patients and those of successes from Groups A (ECMO) and B(CMT), respectively. For the Boston study, $n_A = 29$, $n_B = 10$, $s_A = 28$, and $s_B = 6$. Now, let z be the standardized difference of the two observed proportions s_A/n_A and s_B/n_B, which is approximately 2.305 for the Boston trial. Then a $100(1-\gamma)\%$ exact unconditional confidence interval for Δ is $(\underline{\Delta}, \overline{\Delta})$, where

$$\underline{\Delta} = \inf_{0\le\Delta\le 1}\left\{\Delta: \left[\max_{p_B} \Pr(Z \ge z)\right] > \gamma_1\right\},$$

$$\overline{\Delta} = \sup_{0\le\Delta\le 1}\left\{\Delta: \left[\max_{p_B} \Pr(Z \le z)\right] > \gamma_2\right\}.$$

Here, the random variable Z is the standardized difference $S_A/N_A - S_B/N_B$, where N_A and N_B are generated from the present design, and S_A and S_B from Bernoulli trials with success probabilities p_A and p_B, respectively. In addition, $\gamma_1 + \gamma_2 = \gamma$, and the range of p_B for the maximization is $(0, 1-\Delta)$ if $\Delta \ge 0$, and $(-\Delta, 1)$ if $\Delta < 0$. With various confidence coefficients, these exact unconditional intervals for the Boston ECMO data are presented in our Table 1. It is interesting to note that the 95% one-sided interval is (0.035,1), which barely makes the ECMO effect significant at the 5% level.

For the Boston ECMO trial, the conditional distribution of S_A given $N_A = n_A$, $N_B = n_B$ and $S_A + S_B = s_A + s_B$ which takes the actual design into consideration is degenerate. Hence, the exact conditional confidence intervals as described in Wei, Smythe, Lin and Park (1990) are not applicable. On the other hand, a conventional way to analyze this type of data is by conditioning on n_A, n_B and $s_A + s_B$ without referring to the design actually used in the trial. Then the conditional confidence intervals for the odds ratio can be easily constructed using the noncentral hypergeometric distribution. These intervals are also reported in our Table 1. Here, the lower bound of the 99% interval is literally equal to the null value 1. Thus, if one ignores the design in the analysis, the degree of significance of the ECMO effect will be exaggerated.

Medical investigators should be encouraged to use a data-dependent treatment allocation rule when the new therapy has a great potential benefit over the conventional one. We are currently trying to convince the investigators of our Cancer Center to use adaptive designs for a couple of their phase 2 trials. The design which we would like to use in a future trial is similar to that of Professor Ware's. We will use, for example, two blocks with size four in the first randomized phase. Then an urn model which reflects the relative merit of the two treatments at this stage is constructed using those eight responses. Starting with this urn, a randomized play-the-winner rule with possible early termination will be used in the second phase.