Comment

Richard Royall

This study is deeply disturbing to me. I believe that no babies should have been randomized to "conventional medical therapy." And given that some were, I believe the decision to use the "randomized consent" procedure, whereby their parents were not informed of that fact, was a grave error. It is impossible to present, in this discussion, all of the ethical and statistical arguments that, in my opinion, fully justify these judgments; I can only offer some observations that I hope will convince the readers of *Statistical Science* to give them serious consideration.

Jim Ware is careful to defend the decision to stop randomizing after 19 patients. His calculations show that "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." He also finds that "the posterior probability that ECMO has a lower mortality rate is at least .90 over a wide range of priors," and concludes that "Given this analysis it is difficult to defend further randomization ethically."

Recall that this is not the first randomized clinical trial of ECMO versus CMT. The Michigan group (Bartlett et al., 1985) reported on a trial using a play-the-winner rule that randomized 10 patients, 9 to ECMO and 1 to CMT. All 9 ECMO patients survived and the CMT patient died. Ware and Epstein (1985) published comments on that trial, finding that "... the results are not completely convincing. Why not? Primarily because only one patient received the standard therapy." They concluded 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."

Thus before the Harvard study began, there were 10 cases available from a carefully conducted randomized controlled trial. The main problem was that there were not enough babies in the control (CMT) group. But were there? Using the same prior probability distributions that Ware does, we find that, before the Harvard study was begun, the posterior probability that ECMO is inferior to conventional therapy, given experience in the 10 randomized patients at Michigan, was only .01 when using the prior that ignores Harvard's recent (before their own trial) experience with conventional therapy and falls to .00003 when one uses the prior based on that experience. The corresponding probabilities that ECMO is superior were .90 and .9996 respectively. In Ware's own words "Given this analysis it is difficult to defend further randomization ethically," but now "further randomization" refers to the *first baby* entered in the Harvard trial. The argument that the Michigan results were not convincing because of insufficient concurrent randomized controls is refuted by the same reasoning and calculations that Jim Ware uses to defend the Harvard group's decision to stop randomizing babies to their own control group.

It might be argued that conditions were different at Harvard from those at Michigan, with differences in patient populations as well as in how the two therapies were defined and applied. Thus, although the probability that conventional treatment is better at Michigan is only .01, one cannot be sure that the same is true at Harvard. Note that this argument would apply *regardless of the size of the control group* in the Michigan study. It would mean that no matter how strong the evidence that ECMO sharply improves survival at Michigan, the Harvard trial would still have been necessary. But then the Harvard evidence would not be applicable at Johns Hopkins where yet *another* randomized trial would be necessary ... and so on at hospitals throughout the land. Moreover, because populations, personnel and procedures change over time, one cannot be sure that last year's results at Harvard are relevant to tomorrow's therapeutic decisions there. Now caution in extrapolating findings from one time or place to another is an important aspect of scientific rigor. Thus the present argument can lead to the conclusion that historical controls, or even nonrandomized concurrent controls, are of little scientific value, that good science demands that we avoid extrapolating and settle for nothing less than concurrent randomized controls, that good science demanded the Harvard study.

But if the argument leads to that conclusion, it also leads to the conclusion that the randomized trial itself, being conducted at a unique time and place, has little value. If we are unwilling to extrapolate, to make the assumptions about similarity, stability and consistency required for the judgment that clinicians at other times and places have something to learn from our

Richard Royall is Professor, Department of Biostatistics, The Johns Hopkins University, 614 N. Wolfe St., Baltimore, Maryland 21205.