remains to be seen. In any case, as a precedent for important confirmatory studies in the future, I believe that 19 patients is just too small a sample size to be recommended. One could ask the question: what therapy would I choose if I had a child suffering form persistent pulmonary hypertension? Well, I would certainly choose ECMO based on the available evidence. However, I would also choose ECMO even if the data were only 7/9 versus 6/10 in its favor. In other words, when your own neck is on the line you always want to choose the treatment that appears to be best. Unfortunately, if everyone is permitted to do this the resulting anarchy would totally undermine the scientific rationale on which the best modern medical research is based. For this reason, emotive questions like the preceding one tend to cloud our reasoning when we debate the merits of randomization.

ADDITIONAL REFERENCES

BEGG, C. B. and BERLIN, J. A. (1988). Publication bias: A problem in interpreting medical data (with discussion). *J. Roy. Statist. Soc. Ser. A* **151** 419–463.

HELLMAN, S. (1979). Randomized clinical trials and the doctor-patient relationship: An ethical dilemma. *Cancer Clin. Trials* **2** 189–193.

PETO, R. (1985). Discussion of the papers by J. A. Bather and P. Armitage. *Internat. Statist. Rev.* **53** 31–34.

Comment

Peter Armitage and D. Stephen Coad

Dr. Ware has performed a valuable service in two particular respects. He has given us a carefully documented case study, tracing the role of the statistician from the interpretation of past data, to the planning of a new investigation and the analysis and presentation of its results. Editors of statistical journals frequently bemoan the paucity of case studies amongst the papers submitted to them. Here is an excellent example of such a study.

More specifically, Dr. Ware has described one of the very few clinical trials using any form of outcome-dependent allocation. Armitage (1985) has drawn attention to the need for more interchange of ideas and experience between theorists and practitioners concerned with this aspect of clinical trial methodology. Dr. Ware's paper is a welcome contribution to the literature, from both a practical and a theoretical viewpoint.

There are several examples in therapeutic medicine of unresolved questions, for which the evidence relies almost entirely on nonrandomized comparisons, but where investigators have, for ethical reasons, been reluctant to initiate randomized trials. It is hard to resist the view, expressed, for instance, by Chalmers, Block and Lee (1972), that randomized studies ought to be initiated at a very early stage of the introduction of new methods (they would say for the first patient). In the wake of the earlier inconclusive trial of ECMO, and the controversy to which it gave rise, the present investigators naturally had to tread cautiously, and their wish to restrict the use of CMT as far as possible is understandable. In a rather similar, and equally controversial, situation recently, the (British) Medical Research Council gave firm backing to an extensive trial of vitamin supplementation for women becoming pregnant after an earlier pregnancy resulting in a neural tube defect, to see whether supplementation reduces the risk of a further affected infant. Some had argued that evidence from nonrandomized studies was sufficient to justify routine use of supplementation. The MRC took the view that firm and reliable evidence was needed, and that a randomized trial, carefully monitored, was justified (Wald and Polani, 1984). Its results are awaited eagerly.

The evidence for the superiority of ECMO over CMT, patchy as it is, seems to us fairly convincing. Our view, though, is heavily affected by the fact that patients in phase 2, all of whom received ECMO, were apparently at higher risk than those in phase 1. The eligibility criterion was tightened to exclude some less severely affected patients, and a higher proportion than in phase 1 were outborn, a characteristic apparently conferring higher risk. Had this feature not been present we should have been only moderately impressed, on the grounds that the comparability of phases 1 and 2 was in doubt and that the evidence from phase 1 was weak.

As regards phase 1, we are skeptical of any analysis that suggests a difference much more significant than is given by the Fisher exact test. The Bayesian probabilities for $p_1 > p_2$ are small, partly because an arbitrary amount of prior (and therefore posterior)

Peter Armitage is Professor of Applied Statistics and D. Stephen Coad is a research student, Department of Statistics, University of Oxford, 1 South Parks Road, Oxford OX1 3TG, England.

probability is placed on $p_1 = p_2$, and partly (in the case of the beta prior) because the prior with a mean of 0.2 is persisted with in spite of the observed survival of 6 patients out of 10.

However, we are rather more concerned about the suitability of the design, not with the benefit of hindsight but in the light of the considerations available at the planning stage. The combination of a potentially small randomized phase 1 with a nonrandomized phase 2 involving 100% allocation to the apparently better treatment was likely to place a great deal of weight on the comparability of patients in the two phases. As we have seen, things turned out well, but the lack of comparability might well have favoured rather than disfavoured ECMO; interpretation would then have been very difficult. A second point is that phase 1 might have ended with very little evidence of a difference: say, 4 deaths out of 10 on CMT and 2 out of 9 on ECMO. We think that there would then have been every reason to continue randomization, and no ethical qualms about doing so.

Much of our reservation about outcome-dependent allocation centres around two points. The first is that the more one arranges to move adaptively towards assignment of the better treatment (and thus reduce the number of patients receiving the inferior treatment), the less efficient the estimate of the treatment effect. This can be shown analytically for certain designs for comparing proportions or normal means, and we have confirmed it by simulation for several more complicated allocation rules. We agree that the objective of precise estimation in a clinical trial must be subservient to ethical constraints. Nevertheless, a trial in which the use of an inferior treatment is minimized but which provides a very imprecise estimate may have less impact on the medical community, and hence be slower to achieve a change in medical practice, than a more conventional equal-allocation trial.

The second reservation is the effect of secular changes in the general level of response in the properties of different assignment rules, an instance of which is the contrast between phases 1 and 2 of the present trial. Our simulations have examined the effect of gradual trends in response. For modest trends, and for assignment rules that change the allocation probabilities gradually, the effects do not seem too serious. But sudden changes in level of response, associated with sudden changes in assignment (as in phase 2 here) may be more catastrophic.

Finally, a remark about sequential methods, which Dr. Ware says "do not use adaptive randomization." This is surely an oversimplification. His own study is sequential, in that the termination point depends on the data. He presumably has in mind standard SPRT or closed designs. It is worth bearing in mind that these stopping rules can be used in conjunction with adaptive randomization, as shown for instance by Flehinger and Louis (1972), Robbins and Siegmund (1974) and Hayre (1979).

ACKNOWLEDGMENT

D. Stephen Coad is supported by the Medical Research Council.

ADDITIONAL REFERENCES

ARMITAGE, P. (1985). The search for optimality in clinical trials. *Internat. Statist. Rev.* **53** 15–24.

CHALMERS, T. C., BLOCK, J. B. and LEE, S. (1972). Controlled studies in clinical cancer research. *New England J. Med.* **287** 75–78.

FLEHINGER, B. J. and LOUIS, T. A. (1972). Sequential medical trials with data-dependent treatment allocation. *Proc. Sixth Berkeley Symp. Math. Statist. Probab.* **4** 43–51. Univ. California Press.

HAYRE, L. S. (1979). Two-population sequential tests with three hypotheses. *Biometrika* **66** 465–474.

ROBBINS, H. and SIEGMUND, D. (1974). Sequential tests involving two populations. *J. Amer. Statist. Assoc.* **69** 132–139.

WALD, N. J. and POLANI, P. E. (1984). Neural-tube defects and vitamins: The need for a randomized clinical trial. *Brit. J. Obstetr. Gynaecol.* **91** 516–523.