problem for the zero-way and one-way eliminations of heterogeneity. He does not discuss the problem of optimal design. Spurrier (1988) presents a method of finding asymptotic "optimal" designs for the one-way elimination of heterogeneity problem based on a variation of asymptotic relative efficiency.

It is interesting to note that at times one gets different "optimal" designs for normal theory and nonparametric analyses. This fact suggests that the question of the normality assumption needs to be addressed at the design stage. I feel that there is much more research to be done in the nonparametric approach to these problems.

ADDITIONAL REFERENCES

DUNNETT, C. W. (1955). A multiple comparison procedure for comparing several treatments with a control. *J. Amer. Statist. Assoc.* **50** 1096–1121.

DUNNETT, C. W. (1964). New tables for multiple comparisons with a control. *Biometrics* **20** 482–491.

NIZAM, A. (1987). Optimal designs when comparing experimental treatments with a control. M.S. thesis, Univ. South Carolina.

SPURRIER, J. D. (1988). Generalizations of Steel's treatments-versus-control multivariate sign test. *J. Amer. Statist. Assoc.* **83** 471–476.

STEEL, R. G. D. (1959a). A multiple comparison sign test: Treatments versus control. *J. Amer. Statist. Assoc.* **54** 767–775.

STEEL, R. G. D. (1959b). A multiple comparison rank sum test: Treatments versus control. *Biometrics* **15** 560–572.

Comment

R. J. Owen

I would like to commend the authors for this broad and valuable review of the literature on optimal designs for comparing treatments with controls.

My first comments refer to the case where there is more than one control. An important situation where this happens is when an experimental treatment is tested against a placebo, the current commercial product and some of the competitors products. In such cases, interest in the different controls may not be the same and it would be useful to have results which take account of this using suitable asymmetric design criteria.

The second point I would like to make concerns the use of prior information. Although those who do not fully accept the Bayesian position often feel uncomfortable with using prior information in the analysis of an experiment, they are less inhibited in its use in choosing a design. Indeed experimenters readily accept that effort should be concentrated where the uncertainty is greatest. The material outlined in Section 7 therefore has potential appeal outside strictly Bayesian circles. In comparing treatments with controls our experiment would often be part of a sequence of several similar experiments. In such cases the information from previous experiments should influence the current design.

When experiments are part of a sequence the choice of design of a single experiment considered in isolation, or even considered posterior to the preceding experiments, is no longer necessarily the appropriate question. Rather it may be better to consider the global design of the whole sequence of experiments. Of course the potential information of one experiment influences the designs of the other experiments. This may be illustrated in the context of continuous designs, one control and one-way elimination of heterogeneity as follows.

Assume the model formulation of Theorem 7.1 and consider two distinct cases: (a) each treatment may appear in each experiment and (b) different experiments can only have the control in common.

In each case let the prior distribution for $(\boldsymbol{\theta}', \boldsymbol{\gamma}')$ take the form stated in Section 7.1 immediately after the error distribution. In case (a) assume that all treatment contrasts with the control $(\theta_1, \theta_2, \cdots \theta_v)$ are exchangeable with respect to their *a priori* dispersion (as in Theorem 7.1). In case (b) this type of exchangeability is also assumed, but now only for the contrasts within an experiment whereas between experiments *a priori* independence of the nonoverlapping sets of treatment contrasts is assumed. This formulation could be appropriate when each experiment is believed to be dealing with a different type of treatment. In both cases consider the criterion of Bayes A-optimality for the sequence of experiments considered as a whole.

In case (a) the overall optimal allocation is given by Theorem 7.1 and its associated algorithm. This still leaves some freedom of choice for individual experiments and it is clear that each experiment may have a design which is exchangeable with respect to the noncontrol treatments. Moreover the first experiment, when considered in isolation, may be taken to be

R. J. Owen is Lecturer, Department of Mathematics, University College of Wales, Aberystwyth, Dyfed, Wales, United Kingdom.

optimal. However, subsequent experiments would not in general be optimal either when considered in isolation or when considered in the light of the information from the preceding experiments.

In case (b) let D_k denote the dispersion matrix of the θ's in the kth experiment conditional on the data of all the experiments. The overall design criterion is therefore to minimize (for K experiments):

$$\sum_{k=1}^{K} \operatorname{tr} D_k.$$

First, imagine fixing the amount of control per block in each experiment and consider optimization with respect to allocation of the noncontrol treatments. It follows that, in view of the form of the prior distribution, Theorem 7.1 applies to each experiment separately and hence the exchangeable allocation of noncontrol treatments is optimal in each experiment. Let $\mathbf{x}_k$ denote the 'x' of Theorem 7.1 in the kth experiment and $D_k(\mathbf{x}_k)$ the value D_k takes with this optimal design so $\operatorname{tr} D_k(\mathbf{x}_k)$ is now its preposterior risk. Hence the overall design problem reduces to

$$\min_{\mathbf{x}_1,\cdots,\mathbf{x}_k} \sum_{k=1}^{K} \operatorname{tr} D_k(\mathbf{x}_k).$$

This criterion would use less control in each experiment than would be optimal in that experiment when considered in isolation.

Note that if the experiments are performed sequentially in time in either case (a) or case (b), then in order to use all the currently available information each experiment needs to be reanalyzed after each subsequent experiment is performed. Observe too that, in either case, missing blocks in some experiments are permitted.

Finally I congratulate the authors on providing a key reference in this important and active research area.

Rejoinder

A. S. Hedayat, Mike Jacroux and Dibyen Majumdar

We wish to thank the discussants for their responses. They have greatly enhanced the article by their thoughtful comments and intriguing questions. Some historical color has been added as well as some new references. We shall briefly address some of the issues which have been raised.

1. CHOICE OF CRITERIA

Several discussants, Bechhofer and Tamhane, Notz, Spurrier and Giovagnoli and Verdinelli have raised the question as to what is the most appropriate optimality criterion to use for the problem being considered here. The alternative criteria suggested can be readily divided into two categories. Bechhofer and Tamhane, Notz and Spurrier suggest that usage of criteria that select designs that maximize the confidence coefficient or that in some sense minimize the size of certain simultaneous confidence regions that can be computed for the $(t_i - t_0)$'s. Giovagnoli and Verdinelli considered some other criteria for point estimates. Before looking at these alternative suggestions, let us recall the optimality criteria we used.

We have followed the classical approach of Kiefer. It is a semiparametric approach, in which we only insist on structures for the first and the second moments of the random variables involved. As for the first moment, we insist on a linear model, and as for the second moment we assume homoscedasticity. We had in mind for the process of selecting a best design the goal of estimating the $(t_i - t_0)$'s with as much precision as possible in the sense of having small variances for the $(\hat{t}_i - \hat{t}_0)$'s. Two of the standard criteria used to accomplish this goal are to select designs that minimize $\sum_{i=1}^{v} \operatorname{var}(\hat{t}_i - \hat{t}_0)$ or minimize the maximal variance of the $(\hat{t}_i - \hat{t}_0)$'s. These criteria are called the A- and the MV-optimality criteria, respectively, and are the criteria upon which we concentrated. So, with only the assumption of a homoscedastic linear model, we are able to control the size of the second moments of the $(\hat{t}_i - \hat{t}_0)$'s in a simple yet meaningful way.

A reservation expressed by Bechhofer and Tamhane and Giovagnoli and Verdinelli concerning usage of the A- and MV-optimality criteria is that these criteria do not take into account the correlations that generally exist between the $(\hat{t}_i - \hat{t}_0)$'s. However, we note that the A- and MV-optimality criteria are closely related in the sense that they will usually select the same design or at least designs that are combinatorially close in structure as being optimal with the MV-optimal designs typically being simpler to identify. It should also be noted that under a given design d, if we let V_d denote the covariance matrix of the $(\hat{t}_i - \hat{t}_0)$'s, then $\sum_{i=1}^{v} \operatorname{var}(\hat{t}_i - \hat{t}_0)$ is equal to the sum of the eigenvalues of V_d. Clearly, these eigenvalues and their