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 associated eigenvectors in general take into account most features of the design including the variance and the covariances of the $(\hat{t}_i - \hat{t}_0)$'s. Thus it seems that, through the spectra of the covariance matrices V_d of the various competing designs, the A-optimality criterion does in fact indirectly take into account the correlations between the $(\hat{t}_i - \hat{t}_0)$'s. We now consider the alternative criteria for selecting a best design suggested by some of the discussants.

1.1 Rectangular Confidence Regions

The criteria of maximizing the joint confidence coefficient (suggested by Bechhofer and Tamhane) or of minimizing the sum of the lengths for simultaneous confidence intervals (suggested by Spurrier) are parametric approaches for selecting an optimal design. These approaches require not only the assumptions stated above for the A- and MV-optimality criteria, but also require an extra assumption concerning the form of the distribution of the random variables involved. The most commonly made distributional assumption in these approaches is that of normality. Of course, one can use a parametric criterion for selecting an optimal design if one is willing to be burdened with the extra distributional assumption. However, we feel that in most situations, it is better to be able to prescribe highly efficient designs based on as few assumptions as possible.

We also feel, along with Spurrier, that A-optimal designs will perform fairly well with respect to simultaneous interval criteria, such as described above. The reason for this being that the formula for any set of simultaneous confidence intervals for the $(t_i - t_0)$'s will almost certainly involve in a critical way the standard errors of the $(\hat{t}_i - \hat{t}_0)$'s. Thus any design which yields a set of simultaneous confidence intervals having a maximal confidence coefficient or which are optimally "narrow" in some sense will also invariably yield small standard errors for the $(\hat{t}_i - \hat{t}_0)$'s which are exactly or at least approximately optimal under the A- or MV-optimality criteria. Nevertheless, it is unfortunate that not much is known about optimal designs for simultaneous confidence intervals. As Notz and Spurrier point out, the main reason is the technical difficulties which are associated with the problem. Clearly, there is a need for more research in this area. Hochberg and Tamhane (1987) is an excellent book in this area, which would be very useful to researchers.

Although the A- and MV-optimality criteria select designs that give us "the most precise" estimates of the $(t_i - t_0)$'s, these criteria do not provide a specific means for answering such questions as to how to rank the test treatments as compared to the control if one has to. There is, unfortunately, no mention of such problems anywhere in the published literature on optimal design theory. We wonder what Jack Kiefer would have done. One approach to this problem, albeit an *ad hoc* solution, which appeals to us involves the computation of a coefficient of variation (CV) for each $\hat{t}_i - \hat{t}_0$. Suppose that t_i's bigger than t_0 are considered to be better than t_0. Then any test treatment which corresponds to a negative CV would be out of competition. Those test treatments having a positive CV have some merit for further consideration. Clearly the $\hat{t}_i - \hat{t}_0$ with the smallest positive CV is the winner. Now, whether to replace the control with the winner is another question that often arises. We feel that other considerations such as overall improvement, replacement cost, etc. should play a primary role in answering this question and in many cases, this question is best answered by the experimenter rather than the statistician.

1.2 Ellipsoidal Criterion

Let us now turn to another question: Why not consider some other criterion from optimal design theory like D- or E-optimality? Our rationale was simple. A-optimality and MV-optimality both possess simple and statistically meaningful interpretations which never made us feel the need for any other criterion. We would welcome examination of other criteria which are statistically meaningful in the context of comparing test treatments with a control. (A description of many popular and standard optimality criteria is given in Hedayat (1981).) Most well-known criteria do not meet this requirement. For example, BIB designs are D-optimal for comparing test treatments with a control and a D-optimal design minimizes the volume of a confidence ellipsoid for the $(t_i - t_0)$'s. Thus the D-optimality criterion treats all contrasts $t_i - t_j$, $i \neq j$, the same with regard to selecting a best design and does not emphasize in any discernible way the contrasts $t_i - t_0$ which are for the problem of comparing test treatments with a control usually considered to be of primary importance. Hence, the D-optimality criterion does not seem to be very statistically meaningful, even though it does "minimize" the variance-covariance matrix of the $(\hat{t}_i - \hat{t}_0)$'s in some sense. The E-optimality criterion also seems to suffer from a lack of a natural (in this context) statistical interpretation, because it minimizes the maximum variance of the estimators of all normalized linear combinations of the $(t_i - t_0)$'s. It is hard to imagine when arbitrary linear combinations of the $(t_i - t_0)$'s would be of interest. Of course, if the experimenter insists, one can still use the D- or E-optimality criteria for selecting an optimal design.

Giovagnoli and Verdinelli propose the J-criterion. This criterion selects designs that minimize

$\text{var}(\sum_{i=1}^{v} \hat{t}_i/v - \hat{t}_0)$ over all designs d. The contrast $\sum_{i=1}^{v} t_i/v - t_0$ is the average of the contrasts $t_1 - t_0, \cdots, t_v - t_0$. But exactly what does this average tell us? In experiments where two sets of treatments $\{1, \cdots, u\}$ and $\{u + 1, \cdots, u + w\}$ are compared, Majumdar (1986) briefly mentioned the criterion: minimize $\text{var}(\sum_{i=1}^{u} \hat{t}_i/u - \sum_{i=1}^{w} \hat{t}_{u+i}/w)$ as a possible means of comparing the two groups of treatments, each taken as a whole. This type of a criterion would make sense in some agricultural experiments where the choice is to use one or the other set of treatments. For example, there can be situations where the user wants to use several varieties simultaneously to guard against the susceptibility of different varieties to different diseases. So, this criterion is relevant when $u - w$ is close to zero. Indeed, if $u = w$, the criterion measures the difference in aggregate yields between the two competing 'packages' of treatments. We wonder if this explanation can be extended to situations where $u = 1$ and w is somewhat larger.

We believe that in any experiment, the optimality criterion that should be used to select a best design is the one most pertinent to the goals of the experiment. This criterion should have a natural and meaningful statistical interpretation and be computationally feasible. If some other approach such as those suggested by the discussants satisfies these conditions, then one should use it. However, in most experimental situations where test treatments are to be compared to a control and where little prior information is available, we recommend usage of the A- or MV-optimality criteria for selecting a best design. The basis for our recommendation is fivefold:

(1) A- and MV-optimality have clear and natural statistical interpretations.
(2) Very few distributional assumptions are needed.
(3) They are in most cases computationally feasible criteria.
(4) Except for (v, b, k), these criteria do not depend on any additional parameters, like the simultaneous confidence level (see Spurrier's comments).
(5) A- and MV-optimal designs tend to be optimal or at least highly efficient under other non-Bayesian criteria such as those suggested by the discussants.

1.3 Bayesian Methods

Several of the discussants mention Bayesian methods and in Section 7 of our article we indicate some of the known results concerning Bayesian approaches to the problem of comparing test treatments with a control. Situations can easily be envisioned where a Bayesian approach might be applicable to the problem being considered here because often times there is information available before an experiment begins with regard to the control or perhaps even the test treatments. We feel that prior information should of course be incorporated into the design of an experiment. However, just how this information should be incorporated into the experimental design is usually not an easily answered question, i.e., it may be somewhat difficult to quantify this information in the form of a prior density on the parameters being estimated. We feel that any assumption made concerning a prior distribution on the parameters must be justifiable in terms of the information available and not be based primarily on mathematical tractability. One place where prior information might be particularly useful is in the determination of sample size. If one has prior information concerning the variability of the control or the test treatments, then one can use this information to optimally allocate experimental units to treatments. For example, Spurrier says that one needs to observe the control more often than the test treatments. Although this statement is often true, there are situations, such as when prior information indicates that observations on the control are much less variable than those on the test treatments, that may allow the experimenter to replicate the control less often than the test treatments and still achieve some specified level of precision.

1.4 Robustness over Criteria

Notz has raised the question of robustness of A-optimal designs under changes of optimality criteria. This is a useful line of research, but one that could prove difficult, in general. From the experimenter's point of view, although robustness over criteria is important, robustness over models probably gets precedence. The experimenter may go with any one single sensible criterion, more easily than a single model. He would rather have the choice of several main dishes and a single dessert, than a single main dish and several desserts. Let us direct our research efforts on the main dishes first—let us give the experimenter model robust designs. As Owen puts it in a different context: "experimenters readily accept that effort should be concentrated where uncertainty is the greatest." A start has been made in Hedayat and Majumdar (1988), but much more work needs to be done.

2. THE ADMISSIBILITY CRITERION OF BECHHOFER AND TAMHANE

Bechhofer and Tamhane make a very keen observation. Some designs in Table 3 of Hedayat and Majumdar (1984) are not admissible according to the criterion of simultaneous confidence intervals. Notz, on the other hand, claims that A-optimal BTIB designs are admissible. (This result is not explicitly stated in Majumdar and Notz (1983), but we believe it is true.) This apparent contradiction is the result of the fact that Table 3 of Hedayat and Majumdar (1984) does not give BTIB designs that are A-optimal among all designs having parameters v, b and k, but rather gives designs that are A-optimal in the subset of BTIB designs having parameters v, b and k. Notz on the other hand is referring to BTIB designs which are A-optimal among all designs. Hindsight tells us that a better title for Table 3 of Hedayat and Majumdar (1984) would have been "A catalog of designs A-best among BTIB designs...."

In Hedayat and Majumdar (1984), we gave Table 3 in the spirit of probing the merits of designs which were optimal among a subset of all designs, in the hope of getting approximately optimal designs for classes where the optimal designs are as yet unknown. We were exploring the method, but not recommending its indiscriminate use. Even though these designs usually seemed to work very well, this was not always the case. In fact, we gave an example in $D(10, 80, 2)$ which is the set of all block designs with $v = 10$, $b = 80$ and $k = 2$, where it is possible to find a non-BTIB design which is at least 24% more efficient than the A-best BTIB design.

Let us consider one of the examples given by Bechhofer and Tamhane. Suppose $k = 2$, $v = 6$ and $b = 30$. The Majumdar and Notz (1983) method for finding an A-optimal design does not work for this class since the optimal $(t, s) = (0, 18)$, but there does not exist a BTIB $(6, 30, 2; 0, 18)$. The A-best design among the available BTIB designs, given by Hedayat and Majumdar (1984), is

$$d(1) = 6d_0,$$

which is six copies of d_0, where

$$d_0 = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 2 & 3 & 4 & 5 & 6 \end{pmatrix}.$$

Bechhofer and Tamhane observe that if one uses

$$d(2) = 2d_0 \cup d_1,$$

which is two copies of d_0 and one of d_1 where d_1 is the BIB design in 15 blocks of size 2 each based on the six test treatments, then $d(1)$ is only 93.75% efficient with respect to $d(2)$. (If d and e are two designs, then the efficiency of d with respect to e can be defined as $\sum_{i=1}^{v} \operatorname{var}(\hat{t}_{ei} - \hat{t}_{e0})/\sum_{i=1}^{v} \operatorname{var}(\hat{t}_{di} - \hat{t}_{d0})$.) Moreover, because $d(2)$ is based on only 27 blocks, we can add any 3 blocks of size 2 to it to make $d(1)$ even more inefficient with respect to this new design. So, which design should we use in $D(6, 30, 2)$? Cheng, Majumdar, Stufken and Türe (1988) suggest two routes to find highly efficient designs, which were outlined in Section 5.2. The following design is "combinatorially close" to a BTIB design with 18 replications of the control:

$$d(3) = 3d_0 \cup d_2,$$

where

$$d_2 = \begin{pmatrix} 1 & 1 & 1 & 1 & 2 & 2 & 2 & 2 & 3 & 3 & 4 & 4 \\ 3 & 4 & 5 & 6 & 3 & 4 & 5 & 6 & 5 & 6 & 5 & 6 \end{pmatrix}.$$

It is easy to see that, for $d = d(3)$,

$$\sigma^{-2} \sum_{i=1}^{v} \operatorname{var}(\hat{t}_{di} - \hat{t}_{d0}) = 1.968.$$

Bechhofer and Tamhane's proposed design $d(2)$ is 87.5% efficient with respect to $d(3)$, whereas the A-best BTIB design $d(1)$ is 82% efficient with respect to $d(3)$. The minimum possible value of the A-criterion (which cannot be attained in $D(6, 30, 2)$), given by equation (5.1)(i), is 1.9487. The (approximate) efficiency of a design d, with respect to an A-optimal design, may be defined to be the ratio

$$E(d) = \sigma^{-2} \sum_{i=1}^{6} \operatorname{var}(\hat{t}_{di} - \hat{t}_{d0})/1.9487.$$

Using this criterion, the efficiencies of the three designs are

$$E(d(1)) = 81\%, \quad E(d(2)) = 87\%, \quad E(d(3)) = 99\%.$$

There seems to be little doubt that $d(3)$ is the design to be used when $v = 6$, $b = 30$ and $k = 2$. It is not a BTIB design, but the slight deficiency it has with respect to "balance" is more than made up by its use of the "optimal" value of the number of replications of the control. A similar analysis can be done for the other two examples of inadmissibility cited by Bechhofer and Tamhane.

We were disappointed to see an example where the A value of the A-best design among BTIB designs increases when b increases, v, k remaining fixed. Inadmissibility of some A-best BTIB designs, according to the criterion of simultaneous confidence intervals, is unfortunate also. These observations do not reflect on the A- or the MV-optimality criteria because these types of situations do not arise for designs that are A- or MV-optimal "among all designs" with fixed (v, b, k). The fact is that in cases such as cited by Bechhofer and Tamhane, if a design which is A-best among BTIB designs only does not perform very well when compared with the (unrestricted) A-optimal designs in their classes, then the same design will also perform poorly according to the criterion of simultaneous confidence intervals. In these classes, we need to look for efficient designs which are not BTIB designs.

A lesson to be learned from these observations is that we have to be very careful when optimizing over only a subset of all designs, like BTIB designs, no matter how natural they may look for the problem at hand. Due to the discreteness of the problem, we might be leaving out highly efficient designs which are slightly unbalanced. This was illustrated by the example we saw a little earlier. These observations also indicate a strong need to study classes of non-BTIB designs, like GDTD's, more carefully. GDTD's include BTIB designs as a special case. Therefore, the best GDTD is more likely to be the best design, or at least very highly efficient, among all designs. This is supported by the tables given in Jacroux (1987b). It would also be worthwhile to study other types of unbalanced designs. Nevertheless, we continue to believe that examples of classes where A-best BTIB designs perform poorly are relatively isolated, especially when k is not too small compared to v. Our belief is strengthened by the results of Stufken (1988). We hope that more research will be done to examine this and to clearly identify classes where BTIB designs are/are not highly efficient.

3. OTHER ISSUES

Owen has given us an interesting account of how to design a sequence of experiments. We agree with him that a good consulting statistician should take advantage of the sequential nature of an experiment when suggesting a design. More research needs to be done in this area.

Owen's other point, also raised by Bechhofer and Tamhane, concerns the unequal importance of several controls in an experiment. Because it is difficult to arrive at a single asymmetric criterion which applies to most experiments of this nature, these problems perhaps are best solved on a case by case basis. The statistician has to work closely with the experimenter to develop a meaningful criterion. He has to be aware that not all demands may be compatible. Clearly, the efficient designs are going to be asymmetric in nature. We hope that tools such as the ones provided in our paper will be useful to such investigators.

The statements S_0, S_1 and S_2 of Giovagnoli and Verdinelli nicely explain the role of symmetry of the optimal BTIB designs. In practice, it is unfortunate, however, that for a great majority of values of the design parameters (v, b, k), the best BTIB design may not be the best design, and, as we have noted earlier, can sometimes even be very inefficient with respect to the optimal design. The next natural class is that of GDTD's. But, as Giovagnoli and Verdinelli note, these designs have not been very well studied as yet. Clearly there is a need for future research to investigate "nice" asymmetric structures. Spurrier observes that an "optimal design under one criterion is generally close to optimal under other criteria." This is, as noted earlier, compatible with our own experience. It would be interesting to see what the nonparametric optimal designs are. His three reasons for the lack of popularity of the simultaneous confidence interval approach are probably correct.

We thank Notz for adding the historical color. Bechhofer and Tamhane, Notz, Giovagnoli and Verdinelli, Spurrier and Owen have all given some excellent suggestions for future research. As for research on the two-way elimination of heterogeneity model proposed by Bechhofer and Tamhane, we would like to mention the article by Ting and Notz (1987a). Here the authors have generalized the concept of BTIB designs to two-way heterogeneity models. They have also obtained optimal designs. Our own on going research indicates that the theory of F-squares can be used to construct optimal designs for the two-way elimination of heterogeneity model. Some examples of optimal designs given in Notz (1985), Ting and Notz (1987a) and Hedayat and Majumdar (1988) are of this type.

Notz inquired about exact Bayes designs using prior knowledge on the control. Recently, some results on this problem have been obtained by Majumdar (1988).

In conclusion, we would once again like to thank the discussants for their stimulating comments. We also thank the Executive Editor, Morris H. DeGroot, for inviting such an excellent panel of discussants.

ADDITIONAL REFERENCES

HEDAYAT, A. (1981). Study of optimality criterion in design of experiments. In *Statistics and Related Topics* (M. Csörgö, D. A. Dawson, J. N. K. Rao and A. K. Md. E. Saleh, eds.) 39–56. North-Holland, Amsterdam.

HOCHBERG, Y. and TAMHANE, A. C. (1987). *Multiple Comparison Procedures*. Wiley, New York.

MAJUMDAR, D. (1988). Optimal block designs for comparing new treatments with a standard treatment. In *Optimal Design and Analysis of Experiments* (Y. Dodge, V. V. Fedorov and H. P. Wynn, eds.) 15–27. North-Holland, Amsterdam.