

SCHOOL OF OPERATIONS RESEARCH  
AND INDUSTRIAL ENGINEERING  
COLLEGE OF ENGINEERING  
CORNELL UNIVERSITY  
ITHACA, NEW YORK 14853

TECHNICAL REPORT NO. 783

February 1988

DISCUSSION OF A PAPER IN  
STATISTICAL SCIENCE BY  
A.S. HEDAYAT, M. JACROUX AND D. MAJUMDAR

by

Robert E. Bechhofer  
and  
Ajit C. Tamhane\*

Approved for public release; distribution unlimited.

Research partially supported by the U.S. Army Research  
Office through the Mathematical Sciences Institute  
of Cornell University.

\*Department of Statistics and Department of Industrial Engineering and Management  
Sciences, Northwestern University, Evanston, Illinois 60201.



DISCUSSION OF THE PAPER  
"Optimal Designs for Comparing Treatments with Controls"  
by A.S. Hedayat, Mike Jacroux and Dibyen Majumdar

by

Robert E. Bechhofer\*

and

Ajit C. Tamhane

Cornell University

Northwestern University

We had little imagined that BTIB designs for comparing test treatments with a control treatment would generate such a wide research interest among the design theorist community when we first proposed this new class of designs in Bechhofer and Tamhane (1981). Naturally we are very pleased and gratified to note the tremendous progress that has been made in the last seven years in the study of these designs and their extensions, with particular emphasis on the optimality question. The authors (HJM), who have been at the forefront of this research, are to be congratulated for providing a fine survey of the results. We thank the Editor for giving us an opportunity to discuss this survey.

The authors focus their attention on A- and MV-optimal designs. Both of these optimality criteria refer to minimizing suitable functions of the variances of the  $\hat{t}_0 - \hat{t}_i$ , but do not take their correlations into account. (We follow the same notation as in the HJM paper.) Thus the optimal designs derived would seem to be appropriate when the results of the experiment are to be reported in terms of the above point estimates accompanied by their estimated standard errors or in terms of separate confidence interval estimates of the  $t_0 - t_i$  ( $i=1, \dots, v$ ). However, the usefulness of the

---

\* This research was partially supported by the U.S. Army Research Office through the Mathematical Sciences Institute of Cornell University.

conclusions drawn from these estimates will generally depend on their simultaneous correctness, and therefore a suitable simultaneous inference (multiple comparisons) procedure is called for; see Hochberg and Tamhane (1987, Chapter 1) for examples. Under normal theory, operating characteristics of simultaneous inference procedures are generally functions of not only the variances of the  $\hat{t}_0 - \hat{t}_i$  but also their correlations. It is true that, for example, A-optimality is equivalent to minimizing the sum of the lengths of the axes of the simultaneous confidence ellipsoid (assuming the usual normal theory model) for the given contrasts of interest. But curiously, D-optimality, which corresponds to minimizing the volume of the confidence ellipsoid, and which does take into account the variances as well as the correlations, is not a useful optimality criterion for the present problem, as the authors rightly point out.

We believe that the use of these standard optimality criteria due to Kiefer (1958) is questionable in experiments involving multiple comparisons of test treatments with a control treatment, because they do not address the typical inferential goals in such experiments. The authors state that "To begin with we need to postulate a model ..." In the same vein, it is also true that, to decide on a design (optimal or efficient), we need to postulate the types of inferences that will be made based on the data collected from the experiment. The authors make a brief reference to this point when they state that "...our primary goal is to determine which among the test treatments might be better than the control..." However, we do not think that this goal necessarily translates into "... to estimate the magnitude of each  $t_0 - t_i$  with as much precision as possible" without reference to how the resulting estimates will be used to determine the apparently better test treatments. In fact, as we explain below, there are

two types of inferential goals commonly used in these experiments, and both involve taking into account the variances of the  $\hat{t}_0 - \hat{t}_i$  as well as their correlations.

Often, in exploratory stages of an investigation there are a large number of new candidate test treatments, and the goal is to screen out those that are inferior to the control treatment; this is the case, e.g., in drug screening. For this goal the subset selection formulation of Gupta and Sobel (1958) would appear to be appropriate. The test treatments that are selected in this initial experiment can then be studied more intensively in the confirmatory stage. Moreover, it usually is required (e.g., by a regulatory agency such as FDA) that the control also be included in this stage. The goal of this second experiment is to obtain more precise estimates of the control versus test treatment contrasts (and also possibly pairwise treatment contrasts). For this goal the simultaneous confidence estimation approach proposed by Dunnett (1955) would appear to be appropriate.

It is important to note that the simultaneous confidence region provided by Dunnett's procedure is "rectangular" in shape, not ellipsoidal. The rectangular confidence region is more useful and relevant in practice for the present problem because (a) it is easier to interpret, and (b) the ellipsoidal confidence region gives much longer confidence intervals when it is projected onto the  $t_0 - t_i$  axes. This is so because an ellipsoidal confidence region is optimal when all linear functions of the  $t_0 - t_i$  for  $i=1, \dots, v$  are a priori of interest; thus it performs conservatively for the specific simple functions, namely, the  $t_0 - t_i$ , which are the only ones of interest in the present problem. The second point to note is that the operating characteristics of the Gupta-Sobel subset selection procedure (e.g., its probability of correct selection) and the Dunnett simultaneous confidence estimation procedure (e.g., its joint coverage probability) depend

on the variances of the  $\hat{t}_0 - \hat{t}_i$ , as well as on their correlations. In fact, the Gupta-Sobel subset selection procedure and the Dunnett procedure for one-sided simultaneous confidence intervals are very closely related, both being based on the same one-sided multivariate Student t percentage point; the two-sided Dunnett's procedure is based on the corresponding two-sided percentage point.

It was with the above background that one of us (Bechhofer) was motivated to study the problem of optimal allocation of observations for 0-way elimination of heterogeneity designs (completely randomized designs). Dunnett (1955) had shown numerically for his rectangular simultaneous confidence region that the  $\sqrt{v}$  allocation rule, i.e.,  $r_{d0} = \sqrt{v} r_{d1}$  and  $r_{d1} = \dots = r_{dv}$ , is approximately optimal (in the sense of maximizing the joint confidence coefficient for a fixed total sample size  $n$ ) for large values of  $n$ , i.e., for large values of the joint confidence coefficient,  $1 - \alpha$ . Bechhofer (1969) also used the criterion of maximizing the joint confidence coefficient for given  $n$ , and derived the optimal allocation (using a continuous approximation to the sample sizes) for one-sided simultaneous confidence intervals of the form  $\{t_0 - t_i \leq \hat{t}_0 - \hat{t}_i + a \ (1 \leq i \leq v)\}$  for specified "allowance"  $a$  and for any value of  $1 - \alpha$ . He also showed that asymptotically (as  $n \rightarrow \infty$ ) the  $\sqrt{v}$  allocation rule is optimal. These results were extended to two-sided simultaneous confidence intervals of the form  $\{t_0 - t_i \in [\hat{t}_0 - \hat{t}_i \pm a] \ (1 \leq i \leq v)\}$  by Bechhofer and Nocturne (1972).

Now the asymptotically optimal  $\sqrt{v}$  allocation rule corresponds to the A- and MV-optimality criteria (if the integer restrictions on the sample sizes are ignored). Therefore these criteria would seem to apply to the simultaneous confidence estimation problem only for large sample sizes (although it is true that the approach to the asymptotically optimal

allocation is fairly rapid as can be seen in the tables given in Bechhofer and Tamhane (1983a)). However, most of the work in optimal designs is concerned with small sample sizes. This is particularly true for the elimination of heterogeneity designs (e.g., block designs) with which the HJM paper is principally concerned. We now turn our discussion to these designs.

In our studies we focused on 1-way elimination of heterogeneity where the block size  $k$  is less than the total number of treatments,  $v+1$ . Based on symmetry considerations, we proposed a new class of so-called BTIB designs, which have the following statistical balance property:

$$\text{var}(\hat{t}_0 - \hat{t}_i) = \tau^2 \sigma^2 \quad (1 \leq i \leq v)$$

and

$$\text{corr}(\hat{t}_0 - \hat{t}_i, \hat{t}_0 - \hat{t}_{i'}) = \rho \quad (i \neq i', 1 \leq i, i' \leq v);$$

here the parameters  $\tau^2$  and  $\rho$  depend on the design and  $\sigma^2$  denotes the common error variance. We called these designs BTIB because they are balanced with respect to the test treatments. This statistical balance property is equivalent to the combinatorial balance property given in Definition 2.2 of the HJM paper (see Theorem 3.1 in Bechhofer and Tamhane (1981)). We next addressed the problem of finding an optimal BTIB design, which for given  $v$  and  $k$ , and for specified standardized "allowance"  $a/\sigma$  and joint (one-sided or two-sided) confidence coefficient  $1-\alpha$ , requires the smallest possible  $b$ . In the search for an optimal design we could eliminate any inadmissible design, which gives a lower joint confidence coefficient for all values of  $a/\sigma$  than another design with a no larger  $b$ . A design that is not inadmissible is called admissible. We characterized inadmissible designs by the following simple rule (see Theorem 5.1 in Bechhofer and Tamhane (1981)) : For given  $k$  and  $v$ , a BTIB design  $d'$  with parameters  $b'$ ,  $\tau'^2$  and  $\rho'$  is inadmissible with respect to another BTIB design  $d$

with parameters  $b$ ,  $\tau^2$  and  $\rho$  if  $b \leq b'$ ,  $\tau^2 \leq \tau'^2$  and  $\rho \geq \rho'$  with at least one strict inequality. This rule is based on the fact that, under normality, the joint confidence coefficient (one-sided or two-sided) is a decreasing function of  $\tau^2$  and an increasing function of  $\rho$ .

Examples of A-optimal designs that are not optimal for our simultaneous confidence interval estimation criterion are easy to find. For example, for

$k=3$ ,  $v=3$  and  $b=18$ , the design consisting of 6 copies of  $\begin{bmatrix} 0 & 0 & 0 \\ 1 & 2 & 3 \end{bmatrix}$  is

given as A-optimal among all designs in Table 1 of Hedayat and Majumdar (1984). From Table OPT1.3.3 of Bechhofer and Tamhane (1985) it is seen that this design is not optimal even in the restricted class of BTIB designs for  $1-\alpha \leq 0.7653$  (but is optimal for larger values of  $1-\alpha$ ). Many more such examples can be found. This is not very surprising, of course, because the two criteria are different. We recognize that different criteria can lead to different optimal designs, and thus it is unfair to judge optimal designs based on one criterion relative to the other. Furthermore, admittedly our criterion is based on the normality assumption, while the authors' criteria are not based on any particular distribution.

However, our admissibility criterion, although also derived from the joint coverage probability calculation under the normality assumption, is much weaker. In other words, if a BTIB design  $d$  requires no more blocks  $b$ , and yet yields no larger variance  $\tau^2\sigma^2$  and no smaller correlation coefficient  $\rho$  than another BTIB design  $d'$  then, in general, the latter should not be used. We were surprised to find that several of the A-optimal BTIB designs given in Hedayat and Majumdar (1984) are inadmissible. In Table 3 of their paper we point out three examples of A-optimal designs that are inadmissible: For  $k=2$  and  $v=3$  let



$$d_0 = \begin{bmatrix} 0 & 0 & 0 \\ 1 & 2 & 3 \end{bmatrix} \text{ and } d_1 = \begin{bmatrix} 1 & 1 & 2 \\ 2 & 3 & 3 \end{bmatrix}.$$

Then for  $b=6$  the design  $2d_0$  with  $\tau^2=1$  and  $\rho=0$  is inadmissible with respect to the design  $d_0 \cup d_1$  (i.e.,  $d_0$  union with  $d_1$ ) which has  $\tau^2=1$  and  $\rho=0.5$ . Similarly for  $b=18$  the design  $5d_0 \cup d_1$  which has  $\tau^2=0.3$  and  $\rho=1/6$  is inadmissible with respect to the design  $4d_0 \cup 2d_1$  which has  $\tau^2=0.3$  and  $\rho=1/3$ . One might say that in each of these two examples both of the competing designs are A-optimal, and Hedayat and Majumdar's algorithm identified the one that unfortunately had the smaller  $\rho$ -value. However, we next give an example where this is not the case. For  $k=2$ ,  $v=6$  and  $b=30$ , the BTIB design consisting of 5 copies of

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

with  $\tau^2=0.4$  and  $\rho=0$  is A-optimal among all BTIB designs. However the design  $2d_0 \cup d_1$  with  $b=27$ ,  $\tau^2=0.3750$  and  $\rho=1/3$  is superior on all three counts, and hence the former is inadmissible; here

$$d_1 = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 2 & 2 & 2 & 2 & 3 & 3 & 3 & 4 & 4 & 5 \\ 2 & 3 & 4 & 5 & 6 & 3 & 4 & 5 & 6 & 4 & 5 & 6 & 5 & 6 & 6 \end{bmatrix}.$$

The design  $2d_0 \cup d_1$  is given as A-optimal for  $b=27$  in Hedayat and Majumdar's (1984) Table 3. But because their algorithm did not compare designs for different  $b$ 's, it failed to note that this design is actually superior (even in terms of the A-optimality criterion) to the A-optimal design for the next higher  $b$ , namely  $b=30$ .

Our admissibility criterion has certain other desirable features, which permit one to restrict consideration for given  $k$  and  $v$  to the so-called minimal complete set of generator designs for constructing any BTIB design for that  $(k,v)$ . The members of the minimal complete set serve as building blocks for larger BTIB designs. Such minimal complete sets were

constructed for selected  $(k,v)$ -combinations by Notz and Tamhane (1983) and Ture (1982). For  $k=2$ ,  $v \geq 2$  and for  $k=3$ ,  $v=3$  it is easy to see that the minimal complete set consists of just two generator designs. In this case the analysis required to determine the admissible and optimal designs is considerably less difficult and is given in Bechhofer and Tamhane (1983b). It may be of interest to note that the result given in equation (3.11) of that paper for characterizing admissible BTIB designs for  $k=2$  is the same as that given in Theorem 3.1 of Hedayat and Majumdar (1984) for characterizing A-optimal BTIB designs.

We conclude our discussion by noting some of the problem areas that need further work. The first on our list is the designs for 2-way elimination of heterogeneity. Much remains to be done in this area, particularly on the problem of constructing "balanced with respect to test treatments row-column designs" (analogous to BTIB designs for 1-way elimination of heterogeneity). The problem of determining the minimal complete set of generator designs for this case is an important one, but quite likely a fairly difficult one. Presumably these designs could be derived from Latin squares, Youden squares and perhaps lattice designs (for large  $v$ ). Some ad-hoc construction methods have been given in a Ph.D. dissertation at Virginia Polytechnic Institute by Rashed (1984).

The next problem on our list is that of finding good designs for comparing test treatments with several controls. As noted by the authors, a beginning has been made in this area of research. In future work it would be desirable to keep some important practical features of the problem in mind. One such feature is that the comparisons with different controls may not be required to be of equal precision. For example, in a clinical trial for a new drug it is not uncommon to include two controls, a placebo and an existing active drug. For regulatory purposes, it often is necessary to

demonstrate the magnitude of the activity of the new drug, and therefore the comparison with the placebo is the more important. It is not necessary to demonstrate to the regulatory agency that the new drug is more effective than the existing drug. But for the purposes of the pharmaceutical company's marketing efforts, in fact, the second comparison is likely to be the more important. This latter comparison would generally be two-sided. Such considerations should be taken into account before determining how to optimally allocate the available experimental resources to different competing test treatments and the controls.

A final brief note concerning nomenclature: We suggest that the word "control" should be used rather than "standard" because the latter usually refers to a known benchmark value. Clearly, in the latter case one cannot use blocking.

We again express our gratitude to the authors for this state-of-the art survey, and to the Editor for giving us an opportunity to comment on it.

## REFERENCES

- Bechhofer, R. E. (1969), "Optimal allocation of observation when comparing several treatments with a control," Multivariate Analysis-II (Ed. P. R. Krishnaiah), New York: Academic Press, 463-473.
- Bechhofer, R. E. and Nocturne, D. J. M. (1972), "Optimal allocation of observations when comparing several treatments with a control, II: 2-sided comparisons," Technometrics, 14, 423-436.
- Bechhofer, R. E. and Tamhane, A. C. (1981), "Incomplete block designs for comparing treatments with a control: General theory," Technometrics, 23, 45-57. Corrigendum, Technometrics, 24, 171.
- Bechhofer, R. E. and Tamhane, A. C. (1983a), "Design of experiments for comparing treatments with a control: Tables of optimal allocations of observations," Technometrics, 25, 87-95.
- Bechhofer, R. E. and Tamhane, A. C. (1983b), "Incomplete block designs for comparing treatments with a control (II): Optimal designs for one-sided comparisons when  $p=2(1)6$ ,  $k=2$  and  $p=3$ ,  $k=3$ ," Sankhyā, Ser. B, 45, 193-224.
- Bechhofer, R. E. and Tamhane, A. C. (1985), "Tables of admissible and optimal balanced treatment incomplete block (BTIB) designs for comparing treatments with a control," Selected Tables in Mathematical Studies, 8, 41-139, Providence: American Mathematical Society.
- Dunnett, C. W. (1955), "A multiple comparison procedure for comparing several treatments with a control," Journal of the American Statistical Association, 50, 1096-1121.
- Gupta, S. S. and Sobel, M. (1958), "On selecting a subset which contains all populations better than a standard," Annals of Mathematical Statistics, 29, 235-244.
- Hedayat, A. S. and Majumdar, D. (1984), "A-optimal incomplete block designs for control-test treatment comparisons," Technometrics, 26, 363-370.
- Hochberg, Y. and Tamhane, A. C. (1987), Multiple Comparison Procedures, New York: John Wiley.
- Kiefer, J. (1958), "On the nonrandomized optimality and randomized nonoptimality of symmetrical designs," Annals of Mathematical Statistics, 29, 675-699.
- Notz, W. I. and Tamhane, A. C. (1983), "Incomplete block (BTIB) designs for comparing treatments with a control: Minimal complete sets of generator designs for  $k = 3$ ,  $p = 3(1)10$ ," Communications in Statistics, Ser. A, (Theory and Methods), 12 (12), 1391-1412.
- Rashed, D. H. (1984), "Designs for multiple comparisons of control versus test treatments," Ph.D. dissertation, Department of Statistics, Virginia Polytechnic Institute and State University, Blacksburg, VA.

Ture, T. E. (1982), "On the construction and optimality of balanced treatment incomplete block designs," Ph.D. dissertation, Department of Statistics, University of California, Berkeley.

