where $\hat{t}_i - \hat{t}_0$ is the best linear unbiased estimate of the elementary treatment-control contrast $t_i - t_0$ and $f$ is a convex, nondecreasing function. It is interesting to speculate, therefore, as to whether or not A-optimal BTIB designs are optimal over a large class of "rectangular" optimality criteria. To answer this, of course, requires coming up with suitable criteria as indicated in my first comment. Similar questions might be raised regarding optimal row-column designs also.

As a third comment, I would like to thank the authors for including material on Bayesian approaches to the design problem. Such approaches seem fairly natural in this setting because often the control treatment is a standard treatment about which we have considerable prior information whereas the test treatments are new and less is known about them. One's prior knowledge about the control should be incorporated into the design and, as one would expect, Bayesian results indicate the effect is to reduce the number of replications of the control. To my knowledge, existing Bayesian results have been obtained by allowing approximate designs and optimal designs are often approximate designs. Although seemingly a hard problem, exact Bayesian design results would be quite interesting. Are the authors aware of any research in this direction?

In summary, Hedayat, Jacroux and Majumdar are to be thanked for a readable and thorough survey article. It is to be hoped that this article will stimulate further research and such research will answer, among other things, the questions I have raised above.

# Comment

**A. Giovagnoli and I. Verdinelli**

This is a very useful survey of many known results on optimal designs of experiments when one of the treatments is a control. It comprises a wide variety of results and it is impossible to comment on each one of them in detail. We shall pick up some general themes.

The first remark is on the choice of the optimality criteria. The title is actually somewhat misleading, because the only optimal designs that are surveyed in it are A- and MV-optimal ones. A-optimality and MV-optimality certainly appear to have very intuitive and appealing statistical interpretations and, according to the authors, are the most widely studied criteria for this type of experimental design. It is a rather disturbing thought, however, that neither of these criteria takes into account the covariances of the estimated treatment-control contrasts.

Besides, other criteria may be relevant in this context. For pilot experiments when the control is taken to be known and the interest lies in testing whether or not the overall effect of the new treatments is appreciable, we may want to contrast the *average* new treatment effect with the old one and minimize $\text{var}(\sum_i \hat{t}_i/v - \hat{t}_0)$, i.e., min var $\sum_i (\hat{t}_i - \hat{t}_0)$, $i = 1, \cdots, v$. This criterion, which can be easily extended to the case of more than just one control, is also mentioned by Majumdar (1986) and it seems appropriate to call it J-optimality because it reduces to minimizing trace$(JPC_d - P')$, with $J$ the $v \times v$ matrix of all ones. In Giovagnoli and Righi (1985) and Notz (1985), it is shown that certain J-optimal designs are also E-optimal, where E-optimality is defined as minimizing the maximum variance of all the estimated contrasts $\sum_i c_i(t_i - t_0)$ with $\sum_i c_i^2 = 1$, and conversely some sufficient conditions for E-optimality turn out to ensure J-optimality too. Thus although E-optimality does not appear to have a very natural statistical interpretation when there is a control, E-optimal plans may also deserve attention in some cases.

Lastly we would like to stress that in the Bayesian approach, due to the (possibly) different prior assumptions on the test treatments and the control, it is no longer true that designs which are D-optimal for inference on treatment-control contrasts, i.e., which minimize the determinant of the posterior covariance matrix of those contrasts, are always D-optimal for any set of contrasts. Thus in this case it is worthwhile to look at D-optimality too.

J-optimality shares with A- and MV-optimality (and also with E- and D-optimality, and others) the property of being invariant under all relabeling of the test treatments which leave the control unchanged. We believe this invariance under a suitable group to be the key to many results on optimal designs, and in

*A. Giovagnoli is Associate Professor, Dipartimento di Scienze Statistiche, Universita degli Studi di Perugia, Via A. Pascoli, 06100 Perugia, Italy. I. Verdinelli is Associate Professor, Dipartimento di Statistica, Prob. e Stat. Appl., Universita di Roma, Piazzale A. Moro 5, 00185 Roma, Italy.*

particular to explain the reason why, as pointed out by the authors, "all the optimal designs given in Sections 2 and 3—and also the Bayesian designs, we may add—possess a high degree of balance in many respects." We shall now illustrate this claim. We shall use the symbol $C_{d(0)}$ for the information matrix of model (1.1), namely $C_{d(0)} = \text{diag}(r_{d0}, r_{d1}, \cdots, r_{dv})$ and $C_{d(1)}$ for that of (1.2), which is called $C_d$ in Section 2.1, whereas $C_{d(2)}$ is defined as in Section 2.2: $C_d$ will stand for any of the above matrices, when either the context is clear or it does not matter which one is meant. Denote by $\tilde{C}_d$ the $v \times v$ matrix obtained from the corresponding $C_d$ by deleting the first row and column so that $\tilde{C}_d^{-1} = PC_d^{-}P'$ is proportional to the covariance matrix of $\hat{t}_i - \hat{t}_0$, $i = 1, \cdots, v$. It is easily seen that all the criteria mentioned so far are functions of $\tilde{C}_d$ which are (i) convex; (ii) nonincreasing with respect to the Loewner ordering of non-negative definite matrices (which from now on will be denoted $\geq_L$, namely $A \geq_L B$ iff $A - B$ is non-negative definite); (iii) invariant under relabeling of the test treatments, i.e., under congruence of $\tilde{C}_d$ by elements of $P_v$, the group of $v \times v$ permutation matrices.

Some general results concerning optimality with respect to criteria satisfying (i), (ii), (iii) (also for more general groups than $P_v$), are given by Giovagnoli, Pukelsheim and Wynn (1987) in the case of continuous (approximate) designs. Roughly speaking, for all criteria as above, "balanced" designs are optimal for a given number $r_0$ of observations on the control, with only the optimal choice of $r_0$ depending on the particular criterion. This seems indeed to be the main unifying idea that, more or less explicitly, runs through the literature on designs for this type of experimental set-up in the exact case too. We should like to expand on this concept by making separate statements for the zero-, one- and two-way elimination of heterogeneity.

## STATEMENT $S_0$

For zero-way elimination of heterogeneity, a design $d^*$ such that the test treatments are "as equireplicated as possible" is optimal for a given $r_0$ with respect to all optimality criteria $\Phi$ satisfying (i), (ii) and (iii). This is also true under a Bayesian model with the normality assumptions of Section 7.0.

To show this, let $\tilde{r}_d = (r_{d1}, \cdots, r_{dv})$ and note that $\tilde{C}_d = \text{diag}(\tilde{r}_d) - n^{-1}\tilde{r}_d\tilde{r}_d'$; if $\tilde{r}_d^*$ is such that $|r_{di}^* - v^{-1}(n - r_0)| < 1$ for all $i = 1, \cdots, v$, then by integer majorization (see Marshall and Olkin, 1979, page 134) for any $\tilde{r}_d$ such that $\sum_i r_{di} = n - r_0$, there are permutation matrices $\Pi_i$ and real numbers $\lambda_i \geq 0$, $\sum_i \lambda_i = 1$ for which $\tilde{r}_d^* = \sum_i \lambda_i \Pi_i \tilde{r}_d$. This implies $\text{diag}(\tilde{r}_d^*) = \sum_i \lambda_i \Pi_i \text{diag}(\tilde{r}_d)\Pi_i'$ and (see Marshall and Olkin, 1979, page 468, Proposition E.7.a) $\tilde{r}_d^*\tilde{r}_d^{*\prime} \leq_L \sum_i \lambda_i \Pi_i \tilde{r}_d \tilde{r}_d' \Pi_i'$ so that $\tilde{C}_d^* = \text{diag}(\tilde{r}_d^*) - n^{-1}\tilde{r}_d^*\tilde{r}_d^{*\prime} \geq_L \sum_i \lambda_i \Pi_i(\text{diag}(\tilde{r}_d) - n^{-1}\tilde{r}_d\tilde{r}_d')\Pi_i' = \sum_i \lambda_i \Pi_i \tilde{C}_d \Pi_i'$. From properties (i), (ii) and (iii) of $\Phi$ it follows that $\Phi(\tilde{C}_d^*) \leq \Phi(\tilde{C}_d)$ for all $d$ so $d^*$ is optimal for a given $r_0$.

The optimal values of $r_0$ for A- and MV-optimality are obtained, respectively, from equation (2.3) and, as indicated in Section 3.0, from $\min_d(1/r_{d0} + 1/p(r_{d0}))$. For J-optimality we must solve $\min_d v^2/r_{d0} + (v - n + r_{d0} + vp(r_{d0}))/p(r_{d0}) + (n - r_{d0} - vp(r_{d0}))/(p(r_{d0}) + 1)$. Thus J-optimal designs will in general be different from A- and MV-optimal ones.

Clearly in the Bayesian case the very same properties of the criteria hold as functions of the posterior covariance matrix of treatment-control contrasts. For zero-way elimination of heterogeneity the posterior information matrix for the treatment-control contrasts is given by $\tilde{C}_d + T^{-1}$, where $T$ is the prior covariance matrix, which under our hypotheses is invariant under congruences by $P_v$. Special cases of such $T$ are obtained from matrix $C_2$ of Section 7.0 and matrix $H$ of Equation 11 in Smith and Verdinelli (1980). Clearly $\tilde{C}_d^* + T^{-1} \geq_L \sum_i \lambda_i \Pi_i(\tilde{C}_d + T^{-1})\Pi_i'$. Optimal values of $r_{d0}$ have been computed, however, only in the continuous case, as pointed out in the present paper.

Our next statement will be based on the following proposition which is a very easy generalization of a well-known result by Kiefer (1975).

## PROPOSITION

Given a class of $v \times v$ matrices $C = \{\tilde{C}_d\}$, if a matrix $\tilde{C}_d^* \in C$ (1) is completely symmetric, i.e., invariant under congruences by $P_v$, (2) maximizes $\text{tr}(J\tilde{C}_d)$ in $C$, (3) maximizes $\text{tr}[(I - v^{-1}J)\tilde{C}_d]$ in $C$, then $\tilde{C}_d^*$ minimizes in $C$ all criteria with properties (i), (ii), (iii).

This is Proposition 3 of Giovagnoli and Wynn (1985b).

## STATEMENT $S_1$

For one-way elimination of heterogeneity, with $b$ blocks of size $k$, a design $d^*$ such that (a) $\tilde{C}_d^*$ is completely symmetric, (b) $n_{d0j}^*$ are "as equal as possible," $1 \leq j \leq b$, and (c) $n_{dij}^*$ are "as equal as possible" for each given $j = 1, \cdots, b$ ($\forall i = 1, \cdots, v$) (if $k \leq v$, this means that $n_{dij} \in \{0, 1\}$ $\forall i \neq 0$), is optimal for a given $r_0$ with respect to all criteria satisfying (i), (ii) and (iii).

The proof of this statement follows from the proposition above, if one computes the trace and the sum of the elements of the matrix $\tilde{C}_d$. Condition (b) implies (2) of the proposition, and (c) implies (3), by minimizing $\sum_j [\sum_i n_{dij}^2 - v^{-1}(k - n_{d0j})^2]$.

Statement $S_1$ is very similar to Theorem 4 of Giovagnoli and Wynn (1985), but note that there optimality criteria were taken to be invariant under congruence of $\tilde{C}_d$ by *orthogonal* transformations (Schur-optimality) which leaves out MV-optimality. $S_1$ explains the A- and MV-optimality of the ABIB- and BTIB-designs (both of the rectangular and of the step type) mentioned in Sections 2.1 and 3.1 of the present article, whose number $r_0$ of observations on the control is "right." Observe that design $d_1$ of Section 5.2 satisfies (b) but not (a) of $S_1$, while the converse holds for $d_2$.

Condition (a), i.e., the complete symmetry of the matrix $\tilde{C}_d$, ensures that $\Phi(\tilde{C}_d^{-1})$ can be easily computed. It may be argued that (a) is not simple to achieve, either in the exact or in the continuous case, and this is where the GDTD's of Theorem 2.2 come in as "second best." For such designs the invariance is under, with obvious notation, the group $P_m \otimes P_q$. We do not know of a theory of invariant matrix orderings developed for this group.

For Bayes designs, whatever the prior covariance of the blocks: For *one-way elimination of heterogeneity*, with $b$ blocks of size $k$, under normality assumptions and with completely symmetric prior covariance matrix for the treatment-control contrasts, a design $d^*$ such that the $n^*_{dij}$'s are all equal, $i = 1, \cdots, v$; $j = 1, \cdots, b$, is optimal for a given $r_0$ with respect to all criteria satisfying (i), (ii) and (iii).

This may be a suitable point to remark that

$$C_{d(1)} = C_{d(0)} - k^{-1}N_d(I - b^{-1}J)N'_d \leq_L C_{d(0)}.$$

If $n_{dij} = r_i/b \ \forall j$ (this is known in design theory as orthogonality of treatments and blocks), then $C_{d(1)} = C_{d(0)}$. Thus a design $d^*$ which is orthogonal and optimal—with respect to any criterion satisfying (ii)—for zero-way elimination of heterogeneity is also optimal for one-way elimination of heterogeneity.

For two-way elimination of heterogeneity, the same argument holds, because $C_{d(2)} = C_{d(1)} - b^{-1}M_d(I - k^{-1}J)M'_d \leq_L C_{d(1)}$. Thus, orthogonality of treatments and rows (mentioned as condition (ii) both in (2.8) and in Section 3.2) leads to optimal designs for one-way elimination of heterogeneity being optimal for two-way elimination of heterogeneity too.

## STATEMENT $S_2$

For *two-way elimination of heterogeneity*, with $k$ rows and $b$ columns, a design $d^*$ such that (a) $\tilde{C}^*_d$ is completely symmetric, (b) $n^*_{d0j}$ are "as equal as possible," $j = 1, \cdots, b$, (c) $n^*_{dij}$ are "as equal as possible" for each given $j = 1, \cdots, b (\forall 1 \leq i \leq v)$ and (d) $m^*_{dih} = r_i/k$ for each $h = 1, \cdots, k$, $i = 1, \cdots, v$, is optimal for a given $r_0$ w.r.t. all criteria satisfying (i), (ii) and (iii).

Theorem 2.1 of Notz (1985) is a special case. The optimal designs of Sections 2.2 and 3.2. satisfy the conditions (a), (b), (c) and (d) of $S_2$.

One further remark may be made about the examples of Section 7. Individual numerical examples sometimes look fairly artificial—like the choice of $n = 36$ and the values 1/6 and 1/18 for the ratios of the prior treatment variance and the prior variance of the control to the error variance in example 7.1. It might have been more interesting for the readers to show the behavior of the Bayes designs for zero-way elimination of heterogeneity for various values of the parameters in the prior distribution. Such comparisons may lead to a better qualitative understanding of the optimal designs obtained. When the prior variance of the control is very small with respect to the error variance, as in Example 7.1, it means that our knowledge about the control is fairly accurate, and the number of the observations on the control will be much smaller than in the classical case when no prior information is available (prior variances going to infinity). On the other hand, making use of the three-stage hierarchical model with vague knowledge at the third stage as in Example 7.2. corresponds to vague knowledge about the control treatment and precise information on the test treatments; this gives an even larger number of observations on the control. Note that the considerations on the limiting behavior of the Bayes A-optimal designs in Smith and Verdinelli (1980) hold true despite an algebraic error in the quartic equation at page 617 of that article.

## ADDITIONAL REFERENCES

GIOVAGNOLI, A. and RIGHI, M. (1985). Construction of J-optimal designs for pilot experiments. *Atti Convegno Giornate di Metodologia Statistica.* Bressanone, Settembre 1985.

GIOVAGNOLI, A., and WYNN, H. P. (1985b). G-majorization with applications to matrix orderings. *Linear Algebra Appl.* **67** 111–135.

GIOVAGNOLI, A., PUKELSHEIM, F. and WYNN, H. P. (1987). Group invariant orderings and experimental designs. *J. Statist. Plann. Inference* **17** 159–171.

MARSHALL, A. W. and OLKIN, I. (1979). *Inequalities: Theory of Majorization and Its Applications.* Academic, New York.