On the Equivalence of Optimality Design Criteria for the Placebo-Treatment Problem

Holger Dette
Ruhr-Universität Bochum
Fakultät für Mathematik
Universitätsstr. 150, D-44780 Bochum, GERMANY

Weng Kee Wong
Department of Biostatistics, University of California,
Los Angeles, CA 90095-1772

Wei Zhu
Department of Applied Mathematics and Statistics
State University of New York, Stony Brook, NY 11794-3600

Abstract

We consider a class of optimality criteria and show that each criterion has its unique and equivalent dual within the class. This property can be used to find a variety of optimal designs, including a class of compound optimal designs and their relationships. As an example, we show that one type of D-optimal design provides analytical formula for a class of compound optimal designs, while its dual, the more traditional criterion, cannot.

Key Words: Compound optimality, D-optimality, design efficiency, minimax optimality.

*All correspondence and reprints should be sent to Weng Kee Wong, Dept. of Biostatistics, UCLA, School of Public Health, 10833 Le Conte Ave, Los Angeles, CA 90095-1772.

1 The Placebo-treatment Comparison Problem

Many experiments involve comparison of several groups receiving different treatments or groups subject to different conditions. Such problems are discussed extensively in the literature; see Fleiss (1986) and, Zhu and Wong (2000), for example. A common example is in a clinical trial, where patients are grouped to receive a different treatment for each group. Sometimes, the comparisons among these groups may be of unequal interest to the researcher. For example, there is a placebo group and there are two other groups, one receiving aggressive treatments and another receiving less aggressive treatments. The primary objective is to compare the performance of the group receiving aggressive treatments relative to the placebo group, and the secondary objective is to compare the performance of the group receiving the less aggressive treatment relative to the placebo group. The design of the study should therefore provide higher precision for the primary comparison than for the secondary comparison.

More generally, consider the situation where there are several objectives in a clinical trial and we need a design that is deemed adequate for all the I objectives, $(I > 2)$. Suppose further that the i^{th} objective can be represented by a functional,$\Phi_i, i = 1, \cdots, I$, and this function is convex over the space of all designs in the design space. The optimal design for the each of the objectives is the one that minimizes the criterion over all other designs. Designs that minimize a function of several design criteria are called compound optimal designs.

In this paper, designs are treated as continuous designs in the sense of Kiefer's (1985). The use of continuous designs simplifies the technical problem and has the advantage that solutions to the problem provide useful guides to design the study more efficiently. We denote an arbitrary design by ξ and denote the proportion of patients assigned to the i^th group by p_i, with the first group always designated as the placebo group. For the design problems at hand, we only need to determine the optimal proportion p_i of patients to be assigned to each of the treatment arms.

The compound optimal design is a compromise design that balances the various competitive objectives in the trial. When the different levels of interest in each of the objectives are specified by the researcher through the values of the weights λ's, the compound optimal deisgn is found by minimizing the functional

$$\Phi\left(\xi \mid \underline{\lambda}\right) = \sum_{i=1}^{I} \lambda_i \Phi_i\left(\xi\right),$$

where $0 \leq \lambda_i \leq 1$ and $\sum_{i=1}^{I} \lambda_i = 1$. Here, each of the weights λ_i is user-selected, with more important objectives being given a larger value for the weight. The above functional is a convex combination of convex functions and so it is also convex. Consequently, the optimal design can be found using technqiues similar to finding an optimal design under a single objective.

To fix ideas, consider the treatment-placebo comparison problem in a clinical trial with one placebo and several treatments. The model is

$$y_{ij} = \beta_i + \varepsilon_{ij},$$

where β_i represents the effect of treatment i, $i = 1, \cdots, K$, and j is the patient indicator, $j = 1, \cdots, N$. We assume that ε_{ij} represents the normally distributed error term with mean zero and constant variance, and the error terms are independently distributed of one another. Our objective i is to estimate $(\beta_{i+1} - \beta_1)$ as precisely as possible, $i = 1, \cdots, K-1$, assuming treatment 1 denotes the placebo group.

Following convention, we measure the worth of a design ξ by its expected Fisher information matrix, $M(\xi)$ (Atkinson and Donev, 1992, p.95). Under our setup, it is straightforward to verify that such matrices are always diagonal. If we let $A_i^T = \begin{pmatrix} 1 & 0 & \cdots & -1 & 0 & \cdots & 0 \end{pmatrix}$, where (-1) is in the $(i+1)\,th$ position, a direct calculation shows

$$\Phi_i\left(\xi\right) = \ln\left|A_i^T M^{-1}\left(\xi\right) A_i\right| = \ln\left(\frac{N}{n_1} + \frac{N}{n_{i+1}}\right) = \ln\left(\frac{1}{p_1} + \frac{1}{p_{i+1}}\right).$$

Here N is the total sample size, n_i is the sample size in the ith group and $n_i/N = p_i$ is the proportion of patients assigned to the ith group, $i = 1, \cdots, K$. In practice, N is pre-determined ; for example, in clinical trials, the researcher should have prior information on the number of patients he or she can realistically recruit into the trial during the given time frame.

When there are different interests in each of the comparisons, we may use different weights for these objectives. For given values of the weights, λ's, let $\Phi\left(\xi \mid \underline{\lambda}\right) = \sum_{i=1}^{K-1} \lambda_i \Phi_i\left(\xi\right)$, where $0 < \lambda_i < 1$, $\forall i$, and

$$\Upsilon\left(\underline{p}, \rho\right) = \Phi\left(\xi \mid \underline{\lambda}\right) + \rho\left(\sum_{i=1}^{K} p_i - 1\right).$$

The optimal design (i.e. the optimal values of p_i^*) can be found by solving the following set of equations:

$$\begin{aligned} \frac{\partial \Upsilon}{\partial p_1} &= \frac{1}{p_1} \cdot \sum_{i=1}^{K-1} \frac{\lambda_i \cdot p_{i+1}}{p_1 + p_{i+1}} + \rho = 0; \\ \frac{\partial \Upsilon}{\partial p_{i+1}} &= \frac{1}{p_{i+1}} \cdot \frac{\lambda_i \cdot p_1}{p_1 + p_{i+1}} + \rho = 0,\ \ i = 1, \cdots, K-1; \end{aligned}$$

and

$$\frac{\partial \Upsilon}{\partial \rho} = \sum_{i=1}^{K} p_i - 1 = 0.$$

Further algebra shows $\rho = -1$ and the above system of equations reduces to

$$\sum_{i=1}^{K-1} \sqrt{p_1^2 + 4\lambda_i p_1} = 2 + \left(K-3\right) p_1 \tag{1}$$

and

$$p_{i+1} = \frac{-p_1 + \sqrt{p_1^2 + 4\lambda_i p_1}}{2},\ \ i = 1, \cdots, K-1. \tag{2}$$

The general analytic solution is not available, but it is interesting to note that when all comparisons are of equal importance, i.e. $\lambda_i = 1/\left(K-1\right)$, $i = 1, \cdots, K-1$, we have

$$p_1^* = \frac{1}{1 + \sqrt{K-1}} \tag{3}$$

and

$$p_i^* = \frac{1}{K - 1 + \sqrt{K-1}} = \frac{1}{\sqrt{K-1}} p_1^*,\ \ i = 2, \cdots, K. \tag{4}$$

This means that when we are equally interested in comparing all $(K-1)$ pairs of placebo and treatment, we should allocate equal number of patients to each of the $(K-1)$ treatments and $\sqrt{K-1}$ times this number of patients to the placebo. The above result is thus a generalization of the well known result given in Fleiss (1986, page 96) for comparing several treatments versus a placebo and there is equal interest in all the comparisons.

2 Another D-optimality Type Criterion

It is instructive to consider an alternative D-optimality criterion given by

$$\tilde{\Phi}(\xi) = \left| A^T M^{-1}(\xi) A \right|,$$

where A is a user-selected semi-positive matrix. The choice for the matrix A depends on the objective of the study. As in D-optimality, we seek a design to minimize this criterion over all designs. Let A_i be as before, and note that we now have

$$\tilde{\Phi}_i(\xi) = \left| A_i^T M^{-1}(\xi) A_i \right| = \frac{N}{n_1} + \frac{N}{n_{i+1}} = \frac{1}{p_1} + \frac{1}{p_{i+1}},$$

where $n_i/N = p_i,\ i = 1, \cdots, K$. Let $\tilde{\Phi}(\xi \mid \underline{\lambda}) = \sum_{i=1}^{K-1} \lambda_i \tilde{\Phi}_i(\xi)$, where $0 < \lambda_i < 1$, $\forall i$, and we have

$$\begin{aligned}
\Upsilon\left(\underline{p}, \rho\right) &= \tilde{\Phi}(\xi \mid \underline{\lambda}) + \rho\left(\sum_{i=1}^{K} p_i - 1\right) \\
&= \sum_{i=1}^{K-1} \lambda_i \left(\frac{1}{p_1} + \frac{1}{p_{i+1}}\right) + \rho\left(\sum_{i=1}^{K} p_i - 1\right) \\
&= \frac{1}{p_1} + \sum_{i=1}^{K-1} \lambda_i \frac{1}{p_{i+1}} + \rho\left(\sum_{i=1}^{K} p_i - 1\right).
\end{aligned}$$

The compound optimal design can be found by solving the following set of equations:

$$\begin{aligned}
\frac{\partial \Upsilon}{\partial p_1} &= -\frac{1}{p_1^2} + \rho = 0; \\
\frac{\partial \Upsilon}{\partial p_{i+1}} &= -\frac{\lambda_i}{p_{i+1}^2} + \rho = 0,\ i = 1, \cdots, K-1;
\end{aligned}$$

and

$$\frac{\partial \Upsilon}{\partial \rho} = \sum_{i=1}^{K} p_i - 1 = 0.$$

Further algebra shows that the above system of equations reduces to

$$p_{i+1} = \sqrt{\lambda_i} p_1, \ i = 1, \cdots, K-1$$

and

$$\sum_{i=1}^{K} p_i - 1 = 0.$$

The general analytic solution is readily obtained as

$$\begin{aligned} p_1 &= \frac{1}{1+\sum_{i=1}^{K-1}\sqrt{\lambda_i}}; \\ p_{i+1} &= \frac{\sqrt{\lambda_i}}{1+\sum_{i=1}^{K-1}\sqrt{\lambda_i}}, \ i = 1, \cdots, K-1. \end{aligned}$$

It is interesting to note that when we are equally interested in all the pairwise comparisons, we set $\lambda_i = 1/(K-1)$, $i = 1, \cdots, K-1$, and obtain the same design as before ,i.e.

$$p_1^* = \frac{1}{1+\sqrt{K-1}} \tag{5}$$

and

$$p_i^* = \frac{1}{\sqrt{K-1}} p_1^*, \ i = 2, \cdots, K.$$

An explanation of this property from a theoretical point of view is given in Corollary 3 of Section 5.

Table 1 shows the two types of D-optimal designs when there are $K = 4$ comparison groups for selected choice of weights. When there is unequal interest in each of the comparisons, the two types of D-optimal designs are different; otherwise, they coincide as the theory just showed. Under both criteria, the D-optimal design assigns more patients to the group deemed more important than the other groups. For instance, the first row of Table 1 shows the comparison between the fourth group and the placebo group is deemed the most important with a weight of 0.7. The proportion of patients assigned to group 4 is 36.7%, which is the highest among the groups $2, 3$ and 4. We also note that in all cases, the placebo group receives the most patients. This makes sense because this is the group most involved in all the comparisons.

Table 1: Two types of D-optimal designs for 3 treatment groups and different weights are used to compare their effects relative to the placebo group (group 1). The proportions in parentheses are those obtained using the criterion without the log.

λ_1	λ_2	λ_3	p_1	p_2	p_3	p_4
0.1	0.2	0.7	0.404 (0.385)	0.083 (0.122)	0.147 (0.172)	0.367 (0.322)
1/3	1/3	1/3	0.366 (0.366)	0.211 (0.211)	0.211 (0.211)	0.211 (0.211)
0.1	0.5	0.4	0.386 (0.377)	0.082 (0.119)	0.287 (0.266)	0.245 (0.238)

3 Convexity of the Two D-optimality Criteria

Atkinson and Donev (1992, pg. 96) pointed out that the reason for adopting the D-optimality criterion $\Phi(\xi) = \ln \left| A^T M^{-1}(\xi) A \right|$ over the criterion $\Phi(\xi) = \left| A^T M^{-1}(\xi) A \right|$ is that "taking the logarithm of the determinant leads to minimization of a convex function, so that any minimum found will certainly be global rather than local." The convexity referred to here is the criterion function as a function of the information matrix $M(\xi)$. From Atkinson and Donev (1992) it follows easily that the criterion $\Phi(\xi) = \ln \left| A^T M^{-1}(\xi) A \right|$ is strictly convex in $\underline{p}$.

We claim that for the above placebo-treatment comparison problem the property of strict convexity is also satisfied for the criterion $\tilde{\Phi}(\xi) = \left| A^T M^{-1}(\xi) A \right|$, and therefore, minimizing the function Φ or $\tilde{\Phi}$ with respect to $\underline{p}$ would provide us with the global minimum in either case.

To verify our claim, we recall that a two times continuously differentiable function f is strictly convex on Ω if and only if $\nabla^2 f(\underline{x})$ is positive definite for all $\underline{x} \in \Omega$ (Hiriart-Urruty and Lemaréchal, 1993). It is easy to see that if we take $f(p_1, p_2, \cdots, p_k) = \sum_{i=1}^{K-1} \lambda_i (1/p_1 + 1/p_{i+1})$, then

$$(\bigtriangledown f)^T = \left(-\frac{1}{p_1^2}, -\frac{\lambda_2}{p_2^2}, \cdots, -\frac{\lambda_{K-1}}{p_{K-1}^2}, \right)$$

and

$$\nabla^2 f = \begin{pmatrix} \frac{2}{p_1^3} & 0 & \cdots & 0 \\ 0 & \frac{\lambda_2}{p_2^3} & \cdots & 0 \\ \vdots & \vdots & \ddots & 0 \\ 0 & 0 & 0 & \frac{\lambda_{K-1}}{p_{K-1}^3} \end{pmatrix}.$$

Therefore $\nabla^2 f$ is positive definite for all $\underline{p}$. With the convexity issue resolved, we are now ready to present a general theorem that is useful for understanding the relationship between the above two classes of compound optimal designs.

4 A Class of Optimality Criteria

Consider optimality criteria of the form

$$\left\{\sum_{i=1}^{K-1} \lambda_i |A_i M^{-1}(\xi) A_i|^{-p}\right\}^{1/p} = \Phi_{p,\lambda}(\xi)$$

where $\lambda = (\lambda_1, \ldots, \lambda_{k-1})$ denotes a weight vector and $p \in [-\infty, 1]$.

The cases $p = -1$ and $p = 0, -\infty$ are of particular interest because they give

$$\begin{aligned}
\Phi_{-1,\lambda}(\xi) &= \{\sum_{i=1}^{K-1} \lambda_i |A_i^T M^{-1}(\xi) A_i|\}^{-1} \\
\Phi_{0,\lambda}(\xi) &= \prod_{i=1}^{K-1} |A_i^T M^{-1}(\xi) A_i|^{-\lambda_i} = \lim_{p\to 0} \Phi_{p,\lambda}(\xi) \\
\Phi_{-\infty,\lambda}(\xi) &= \min_{i=1}^{K-1} (A_i^T M^{-1}(\xi) A_i)^{-1} = \lim_{p\to -\infty} \Phi_{p,\lambda}(\xi).
\end{aligned}$$

Note that the cases $p = 0$ and $p = 1$ correspond to the criteria Φ and $\tilde{\Phi}$ discussed in Section 2 and 3. We now provide a result that shows optimal designs constructed under this class of optimality criteria have a dual relationship. Specifically, every optimal design found with respect to a criterion in this class of optimality criteria is also simultaneously optimal under another criterion of the class provided the weight vector is properly chosen.

Theorem 1: *Assume that $p \in (-\infty, 1]$.*

(1) *Let ξ^* denote a design that maximizes $\Phi_{p,\lambda}$ for the weight vector $\lambda = (\lambda_1, \ldots, \lambda_{K-1})$. Then for any $q \in (-\infty, 1]$ the design ξ^* also maximizes $\Phi_{q,\mu}$, where the weight vector $\mu = (\mu_1, \ldots, \mu_{K-1})$ is given by*

$$\mu_j = \frac{\lambda_j(\frac{1}{p_1} + \frac{1}{p_{j+1}})^{q-p}}{\sum_{i=1}^{K-1} \lambda_i(\frac{1}{p_1} + \frac{1}{p_{i+1}})^{q-p}} \quad j = 1, \ldots, K-1$$

and $p_i = n_i/N$ denotes the proportion of total observations allocated by the design ξ^* to treatment $i (i = 1, \ldots, K-1)$.

(2) *If ξ^* maximizes $\Phi_{-1,\lambda}$ for the weight vector $\lambda = (\lambda_1, \ldots, \lambda_{K-1})$ then ξ^* maximizes $\Phi_{0,\mu}$ for the weight vector $\mu = (\mu_1, \ldots, \mu_{K-1})$, where*

$$\mu_j = \frac{\lambda_j + \sqrt{\lambda_j}}{1 + \sum_{i=1}^{K-1} \sqrt{\lambda_i}} \quad j = 1, \ldots, K-1.$$

Proof. Part (1) follow immediately from Theorem 2.4 in Dette (1993) and a straight forward calculation of the information matrix $M(\xi^*)$ [see Zhu and Wong (2000)]. For part (2) we note that the optimal weights for maximizing $\Phi_{-1,\lambda}$ can be obtained directly using Lagrange's multipliers. The optimal weights for the placebo group and the treatment groups are respectively given by

$$\begin{aligned} p_1 &= \frac{1}{1 + \sum_{i=1}^{K-1} \sqrt{\lambda_i}} \\ &\textit{and} \\ p_{i+1} &= \frac{\sqrt{\lambda_i}}{1 + \sum_{i=1}^{K-1} \sqrt{\lambda_i}} \quad i = 1, \ldots, K-1. \end{aligned}$$

This implies

$$\frac{1}{p_i} + \frac{1}{p_{i+1}} = (1 + \sum_{i=1}^{K-1} \sqrt{\lambda_i})(1 + \frac{1}{\sqrt{\lambda_i}})$$

and the assertion now follows from part (1) for $q = 0$ and $p = -1$.

The next result concerns a maximin type of criterion. Maximin or minimax design criteria are popular and have been studied in the literature since 1950 ; some recent work includes Wong (1992) and Dette (1993). This design criterion is particularly useful if we wish to design a study to minimize the maximal variance of all the estimated contrasts. This criterion is also used in situations where it is roughly known in advance that a set of contrasts may be of interest, but which one of the contrasts will be of ultimate interest is not known until the study is completed. Clearly, the maximization of this criterion

$$\Phi_{-\infty}(\xi) = \min_{i=1}^{K-1} (A_i^T M^{-1}(\xi) A_i)^{-1}$$

is equivalent to minimizing

$$1/\Phi_{-\infty}(\xi) = \max_{i=1}^{K-1}(A_i^T M^{-1}(\xi) A_i),$$

i.e. by maximizing $\Phi_{-\infty}$, we minimize the worst possible variance.

Theorem 2. Suppose we wish to find a design that maximizes $\Phi_{-\infty}$. The optimal proportion of patients assigned to the placebo group and the treatment groups are respectively given by

$$\begin{aligned} p_1^* &= \frac{1}{1+\sqrt{K-1}} \\ p_{i+1}^* &= \frac{1}{\sqrt{K-1}} p_1^* \quad i = 1, \ldots, = K-1. \end{aligned}$$

Proof. The equivalence theorem for the maximin criterion can be derived as follows. First, define the set

$$\mathcal{N}(\xi^*) = \{j \in \{1, \ldots, K-1\} \mid (A_j^T M^{-1}(\xi^*) A_j)^{-1} = \Phi_{-\infty}(\xi^*)\}.$$

Using standard maximin arguments [see Pukelsheim (1993)], it can be shown that ξ^* maximizes $\Phi_{-\infty}$ if and only if there exists nonnegative weights $\alpha_1, \ldots, \alpha_{K-1}$ such that

$$\begin{aligned} \sum_{i=1}^{K-1} \alpha_i &= 1 \\ \alpha_i &= 0 \quad \text{if} \quad i \notin \mathcal{N}(\xi^*) \end{aligned}$$

and the inequality

$$\sum_{\ell=1}^{K-1} \alpha_\ell \frac{(A_\ell^T M^{-1}(\xi) x)^T}{A_\ell^T M^{-1}(\xi) A_\ell} \leq 1$$

holds for all $x \in \{(1, 0, \ldots, 0)^T, (0, 1, 0, \ldots, 0)^T, \ldots, (0, \ldots, 0, 1)^T\}$.

The assertion now follows by a straightforward calculation using the design specified in Theorem 2 and the weights $\alpha_i = 1/(K-1)$. We note that for this particular choice, $\mathcal{N}(\xi^*) = \{1, \ldots, K-1\}$.

Corollary 3: *For any $p \in [-\infty, 1]$ the design specified by Theorem 2 is Φ_{p,λ^*} optimal, where λ^* denotes the uniform weight vector, i.e. $\lambda^* = \{\frac{1}{K-1}, \ldots \frac{1}{K-1}\}$.*

Proof. The proof has been established for $p = -\infty$ [Theorem 2] $p = 0$ and $p = -1$ [Zhu and Wong (2000)]. The remaining cases follow from the first part of Theorem 2 [$p = 0$] for any $q \in (-\infty, 1]$.

5 Summary

It is generally difficult to construct an optimal design when there are several competing objectives in the study. Analytical solution of the optimal design is hardly available. We show in this paper that a D-optimality type of criterion always yields closed form formulae for the optimal designs in the placebo-treatment types of problems. The design criterion can also be usefully embedded into a broader class of criteria and dual relationships among optimal designs are presented. This is especially useful because optimal design under one criterion can now be directly deduced from optimal design found under another optimality criterion. In particular, we can easily find optimal design for a a more complicated criterion using an optimal design found under a simpler criterion such as when $p = 0$. We also consider the special role of the minimax criterion and show that the minimax optimal designs are also optimal with respect to all criteria in the given class.

The choice of an optimality criterion or a set of criteria to work with is problem dependent and usually the researcher has a couple of options. For instance, if interest is centered on estimation, A or D-optimality criteria is frequently used. In practice, it is advisable that the researcher finds several reasonable optimal desgins for his or her problems and compare their sensitivities to model assumptions and robustness properties under a range of criteria. Our paper proposes a class of optimal designs for comparing the placebo group and several treatment groups using a class of optimality crtieria and show that the optimal design has several desirable properties. First, the placebo group always has the highest proportion of patients; this is reasonable because the placebo group is the most used group in the set of comparisons. Second, the treatment groups that are heavier weighted require larger sample sizes, and third, the optimal design can be analytically described. It should also be noted that because of the first and second properties, greater precision is ensured for the more important comparisons.

Acknowledgments. The research of Wong is partially supported by a NIH research grant R29 AR44177-01A1. The work of H. Dette was supported by the Deutsche Forschungsgemeinschaft (DE 502/14-1 and SFB 475, Komplexitätsreduktion in multivariaten Datenstrukturen, Teilprojekt A2).

References

Atkinson, A. C. and Donev, A. N. (1992) Optimum Experimental Designs. Clarendon press, Oxford.

Dette, H. (1993). A New Interpretation of Optimality for E-optimal Designs in Linear Regression Models. Metrika 40, 37-50.

Fleiss, J. L. (1986). The Design and Analysis of Clinical Experiments. Wiley, New York.

Hiriart-Urruty and Lemaréchal (1993). Convex Analysis. Springer-Verlag.

Kiefer, J. (1985). Jack Carl Kiefer Collected Papers III: Design of Experiments. Springer-Verlag. New York.

Pukelsheim, F. (1993). Optimal Design of Experiments. Wiley, New York.

Wong, W. K. (1992). A Unified Approach to the Construction of Mini-Max Designs. Biometrika, 79(3), 611-620.

Zhu, W. and Wong, W. K. (2000). Optimal Treatment Allocation in Comparative Biomedical Studies. Statistics in Medicine, 19(5), 639-648.