Published in final edited form as:

Stat Med. 2008 September 30; 27(22): . doi:10.1002/sim.3318.

OPTIMUM TREATMENT ALLOCATION RULES UNDER A VARIANCE HETEROGENEITY MODEL

Weng Kee Wong^{1,*} and Wei Zhu²

¹Department of Biostatistics, School of Public Health, University of California at Los Angeles, California 90095, USA

²Department of Applied Mathematics and Statistics, State University of New York, Stony Brook, New York 11794, USA

SUMMARY

We provide optimal treatment allocation schemes when the outcome variance varies across the treatment groups and our objectives are to estimate treatment effects with equal or unequal interest. Unlike other optimal designs, such as A-optimal designs, the proposed designs can be found without an iterative scheme. We evaluate robustness properties of the optimal designs to mis-specification in the expected variance from each group and identify situations when popular allocation schemes have poor efficiencies. An application to design a randomized rheumatoid arthritis trial is discussed, along with a potential application to design a cancer screening trial when the main outcome is a continuous variable.

Keywords

continuous design; D_A -optimal design; efficiency; multiple-objective optimal design

1. INTRODUCTION

Our goal is to design a randomized trial to compare several treatments when the variance of the continuous outcome varies across groups. In addition, we want the design to incorporate the possibly unequal interests in the set of comparisons. Design issues are of paramount importance because inappropriate allocation of patients to treatment groups results in inaccurate estimates of treatment effects and waste of resources.

When there are several treatment groups, the optimal allocation scheme depends on the objective or objectives of the study. Most of the optimal allocation schemes are developed under a single design criterion and only for the homoscedastic linear model where the variance of the response is assumed constant across groups. For example, the equal allocation rule is optimal if the only aim in the study is to do a pairwise comparison of all the groups [1] and the response variability is constant across the groups. When this assumption is untenable, we have a heteroscedastic model and as we will show here, unequal size allocation can be an efficient strategy.

Our approach is to apply optimal design theory for continuous designs pioneered by Kiefer and voluminous documented in [2]. Continuous designs are simply defined by the

proportion of subjects assigned to each treatment group after the sample size is pre-determined by cost or other considerations. An early application of continuous design theory to clinical trial is to finding optimal rules for patient assignment in a sequential trial [3]. Interest continues to this day to apply optimal design ideas to find allocation schemes in different settings, that includes optimal design for study with a placebo and several treatment groups [4,5,6]. In what is to follow, we apply optimal design theory to design a randomized trial whose objectives are to compare a continuous outcome from treatment groups, with possibly different emphasis in each comparison, and the variance of the outcome varies across the groups.

Section 2 discusses D_A -optimal designs and we provide the formula for the optimal treatment allocation scheme in Section 3. General guidelines for assigning patients to various treatment groups when there is heterogeneity in the response across groups are given in Section 4. We discuss robustness properties of the D_A -optimal design to model assumptions in Section 5 and in Section 6 we show our approach applies to other optimality criteria as well. We conclude in Section 7 with an application to design a rheumatoid arthritis study and also indicate a potential application to design a cancer prevention trial.

2. D_A -OPTIMAL DESIGNS

Throughout we assume the model is given by $y = f^T(x) \beta + \epsilon$ and ϵ is normally distributed with mean zero and variance $\sigma^2(x) = \sigma^2 / w(x)$. The vector of parameters β are the treatment effects. The positive function $w(x)$ is assumed known and is called the efficiency function in the optimal design literature. In our setup, the value of $w(x)$ for the i^{th} treatment group is denoted by w_i for simplicity, and w_i is inversely proportional to the variance of the outcome from the i^{th} group, $i = 1, 2, \dots, K$. For the i^{th} group, $f(x_i)$ is the zero vector with the 0 in the i^{th} position replaced by 1.

All our designs are continuous and are represented by the pair (x_i, p_i) , $i = 1, \dots, K$ where x_i is the i^{th} group and p_i denotes the proportion of subjects to be selected from the i^{th} group. Following convention, we measure the worth of the design by its information matrix

defined by $M(\xi) = \sum_{i=1}^K w(x_i) \cdot p_i f(x_i) f^T(x_i)$.

A typical goal is to estimate all pairwise comparisons and these objectives can be expressed as linear combinations of β . We first choose an appropriate matrix A and formulate our objectives for estimating $A^T \beta$. Following Atkinson's work [2,7], a suitable design criterion is D_A -optimality given by $\phi(\xi) = \log |A^T M^{-1}(\xi) A|$. This function, as are all criteria in this work, is a convex function of the information matrix [1] and a design that minimizes this criterion over the set of all designs is D_A -optimal. Such an optimal design ξ^* provides the smallest volume for the confidence ellipsoid based on our estimates of $A^T \beta$. A particularly appealing feature of our optimal designs is that they can be found easily without using an iterative scheme. This implies properties for our designs can be much easily studied than the popular but more elusive A-optimal designs frequently used for such problems.

The D_A -efficiency of another design ξ is defined by:

$$E(\xi) = \left\{ \frac{|A^T M^{-1}(\xi^*) A|}{|A^T M^{-1}(\xi) A|} \right\}^{1/s},$$

where s is the rank of the matrix A . High values of this ratio suggests that the design ξ has similar performance as the optimal design. If this value is equal to one half, for example, this

means that twice as many subjects are required to obtain equally precise estimates provided from the optimal design.

For the chosen matrix A of rank, say s , let $B = A \{A^T M^{-1} A\}^{-1} A^T$ and let $d_A(x, \xi) = w(x) \cdot f^T(x) M^{-1}(\xi) B M^{-1}(\xi) f(x)$. A powerful result from optimal design theory, oftentimes referred to as the general equivalence theorem (GET), states that ξ^* is D_A -optimal on the design space if and only if one of the following two conditions hold:

- i. $\max_x d_A(x, \xi^*) = s$;
- ii. $d_A(x, \xi^*) = s$ if x is a support point of the design ξ^* .

Technical details are available in optimal design monographs [1,8]. The practical usefulness of the above result is that we could guess the optimal design and use (i) to verify if it is optimal. For simpler problems, one can use conditions (i) and (ii) to determine the optimal design analytically. Similar equivalence conditions can be derived for other convex objective functionals.

3. AN OPTIMUM TREATMENT ALLOCATION SCHEME

Suppose we wish to do a pairwise comparison of all K treatments. One choice for the $K \times (K-1)$ matrix A is

$$A^T = \begin{bmatrix} 1 & -1 & 0 & \cdots & 0 \\ 1 & 0 & -1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & 0 & 0 & \cdots & -1 \end{bmatrix}.$$

In the appendix, we show that for any design ξ , we have

$$d_A(x_i, \xi) = \frac{1}{p_i} - \frac{w_i}{\sum_{i=1}^K w_i p_i}, i=1, \dots, K.$$

A direct calculation shows if one imposes conditions (i) and (ii) of GET required for an optimal design with $s = K-1$, we arrive at the optimal allocation rule:

An optimal rule: The optimal proportions p_i 's in the D_A -optimal design ξ^* for estimating all pair-wise mean differences in a heteroscedastic model are obtained by solving the equations:

$$[1 - (K-1)p_1] \sum_{i=2}^K \frac{1}{1/p_1 + (K-1)(w_1/w_i - 1)} - (K-1)p_1^2 = 0 \quad (1)$$

and

$$p_i = \frac{w_1}{w_i/p_1 + (K-1)(w_1 - w_i)}, i=2, \dots, K, \quad (2)$$

where $0 \leq p_i \leq 1$ and $\sum_{i=1}^K p_i = 1$. Furthermore, $0 \leq p_i \leq \frac{1}{K-1}$, $i=1, \dots, K$.

This rule has a few interesting properties. We note that (1). $p_i = 0$, when $w_i \rightarrow \infty$, and $w_j/w_i \rightarrow 0$, $j \neq i$ (i.e. the variance of the response from the i^{th} treatment is small relative to all

other variances); (2). $p_i = 1/(K-1)$, when $w_i = 0$, and $w_j/w_i \rightarrow \infty$, $j \neq i$ (i.e: the variance for treatment i is large relative to all other treatments).

When we have only two treatment groups, i.e. $K = 2$, the D_A -optimal design p^* is defined by $p_1/p_2 = \sqrt{w_2/w_1}$. This result is well known and is obtained by substituting $K = 2$ into equation (1) and using the relationship $p_1 + p_2 = 1$, see [9], for example. Note that a naive generalization to K groups is to assign

$$p_i^{na} = \frac{\sqrt{1/w_i}}{\sum_{i=1}^K \sqrt{1/w_i}}, i=1, \dots, K.$$

proportion of subjects to the i^{th} group. However, this is not the optimal rule. For instance, if $K = 3$; $w_1 = w_2 = 1$ and $w_3 = 128$, the naive rule assigns roughly 0.075, 0.075 and 0.850 to the three groups respectively, but the corresponding optimal proportions are 0.252, 0.252 and 0.496. A popular allocation scheme is to assign the same number of subjects to each group. Such designs are generally not D_A -optimal because $p_i = 1/K$ is not a solution to equations (1) and (2) unless $w_j = w$. In the rest of the paper, we denote proportions given by the naive and equal allocation rules by p_i^{na} and p_i^{eq} respectively.

4. A RULE OF THUMB

It is helpful to have a rule of thumb for allocating subjects to different groups when the outcome has different variances from the groups. We begin by first assuming that there are three groups and the three rules are: the D_A -optimal design, the naive rule and the equal allocation rule. Four scenarios are considered - scenario 1: the three groups have equal variance; scenario 2: two groups have equal variance and the third group has a larger variance; scenario 3: two groups have equal variance and the third has a smaller variance and scenario 4: all three groups have unequal variances. The results are summarized below.

Scenario 1: The three groups have equal variances; i.e. $w_1 = w_2 = w_3 = 1$. In this case, the heteroscedastic linear model becomes the homoscedastic linear model and the three rules coincide with each other.

Scenario 2: Two groups have equal variance and the third has a larger variance, i.e. $w_1 = w_2 = 1$, and $w_3 = w < 1$. The optimal proportions for the first two groups are

$$p = \frac{1 - 4w + \sqrt{1+8w}}{8(1-w)}, 0 \leq w < 1.$$

The proportion p is a monotone increasing function of w , and is bounded between 1/4 and 1/3. Note that $p = 1/3$ yields the D_A -optimal design when w tends to unity discussed in scenario 1. Selected results are given in Table I, where e^{eq} and e^{na} represent the efficiencies of the equal allocation rule and the naive rule relative to the D_A -optimal design. Table I shows the equal allocation rule is more than 94% efficient regardless how different the third variance is from the two equal variances. The naive rule is not robust to misspecification of the third variance because its efficiency e^{na} can drop to below 80%. A reason is that the naive rule allocates almost all subjects to the third group when its variance is large. From the optimal rule, optimal proportions of the D_A -optimal design cannot exceed $1/(K-1)$, in this case, 1/2; therefore, the naive rule cannot be a close approximation to the optimal design when the third variance becomes very large.

Scenario 3: Two groups have equal variance and the third has a smaller variance, i.e. $w_1 = w_2 = 1$, and $w_3 = w > 1$. A similar calculation shows that the optimal proportion for the first two groups is the same as in Scenario 2 except the range of validity of w is now larger than unity. One observes that as w increases, the proportion p to be assigned to the first two groups increases from $1/3$ to $1/2$. The naive and equal allocation designs and their efficiencies for this setup are listed in Table II.

In Scenario 3, the naive rule is doing extremely well because e^{na} is always greater than 99%. The equal allocation rule is not as efficient but e^{eq} is usually not lower than 70%. This suggests that the equal allocation rule is reasonably robust. This is because the third treatment variance is smaller than the first two, and by the optimal rule, a smaller proportion of subjects is allocated to the third group. As the variance approaches 0, so does the proportion of subjects allocated to that group. The naive rule does take the variances into considerations, and it converges to the optimal rule as the third group's variance approaches 0. The equal allocation rule does not adjust to the changes in the variances and deviates from the optimal rule as the third variance goes to 0.

Scenario 4: All three groups have unequal variances. Some representative results are listed in Table III.

Our results here and additional work not reported herein suggest that generally the naive rule is efficient as long as $p_i^{na} \leq 1.2/(K-1)$; in this case with $K=3$, this means that $p_i^{na} \leq 0.6$, i . Recall again that the proportions for the D_A -optimal designs do not exceed $1/(K-1)$ and so the proportions by the naive rule must stay within about 1.2 times the upper bound of the optimal proportions in order to be good approximations to the latter. The equal allocation rule seems generally more robust.

We also studied the case when there are $K=4$ groups and a sample of our results are displayed in Table IV. Results in this table, along with additional results for more than 4 groups not included here suggest that if $(K-1)$ variances are equal and the remaining one is larger, the equal allocation rule is more robust; on the other hand, if $(K-1)$ variances are equal and the remaining one is smaller, then the naive rule is generally more efficient. In all other situations, when $p_i^{na} \leq 1.2/(K-1)$, i , the naive rule is generally more efficient; otherwise, the equal allocation rule is more robust. Of course, if all K variances are equal, the three rules coincide. We summarize our results as follows:

1. If all K variances are equal, use the equal allocation rule;
2. If $(K-1)$ variances are about equal and the remaining one is smaller (larger), then the naive (equal allocation) rule is generally more efficient;
3. In all other situations, when $p_i^{na} \leq 1.2/(K-1)$, i , use the naive rule; otherwise, use the equal allocation rule.

These rules may serve as guidelines. Analytical formulae for the optimal proportions can be obtained using Mathematica or Maple for selected design problems. It must be emphasized that the optimal rules apply only if we know the objectives of the trial, the correct form of the linear model and we are confident that preliminary estimates for the variances are reliable.

5. ROBUSTNESS OF D_A -OPTIMAL DESIGNS TO ASSUMPTIONS ON VARIANCES

In practice, information on response variances may either be unavailable or unreliable. Before a design is implemented, it is an excellent practice to check if the design is robust to

model assumptions. This issue cannot be overemphasized and has been a constant theme in many design papers, see for example [10,11,12]. It is therefore important to evaluate if the proposed D_A -optimal designs are robust when the assumptions on the variances are inaccurate.

Let us assume that the true variance for group i is $\sigma^2(x_i) = \sigma^2/w_i$, but we wrongly assume it to be $\sigma^2(x_i) = \sigma^2/w_i$, $i = 1, \dots, K$. The “ D_A -optimal design” is derived under this false assumption. We say this design is robust to the assumptions on variances if its design efficiency relative to the true D_A -optimal design derived using the correct variances remains reasonably high. Table V provides an example where the true weights are (1, 1/2, 1/16). The second column lists wrongly assumed weights, and the design efficiency of the “ D_A -optimal design” derived under each false assumption is displayed in column 3. As our guess deviates further from the true values of the variances, the efficiency decreases, but remains reasonably high. In an extreme case, when we have “inverted” assumptions on the variance structures as in the last row in Table V, we notice that even in this situation, the “ D_A -optimal design” is still more than 80% efficient.

In summary, results in the table suggest that at least in the cases considered “ D_A -optimal design” appear quite robust to wrong assumptions of variances in the different groups. It is unlikely that broad conclusions can be drawn for a general setup and so in practice, one would do well to investigate robustness properties of the optimal design to various model mis-specification before its implementation.

6. ADDITIONAL OPTIMALITY DESIGN CRITERIA

Our proposed design strategy is flexible and can be used to find other types of optimal designs. For instance, suppose we wish to find a A-optimal design frequently used for such a setting as well. Such an optimal design minimizes the sum of variances of the estimated parameters of interest [1,2]. The optimization problem can be solved directly using Lagrange multipliers and optimality of the design can be verified using similar checking conditions described near the end of Section 2.

Suppose there are K treatments and we wish to assign an optimal number n_t subjects to treatment group t^{th} , $t = 1, \dots, K$. The response of the i^{th} subject in the t group is given by

$$y_{ti} = \tau_t + \frac{e_{ti}}{\sqrt{w_t}}, i = 1, \dots, n_t, t = 1, \dots, K,$$

where e_{ti} 's are independent random errors each with mean 0 and variance σ^2 . The estimator for the treatment t effect τ_t is $\hat{\tau}_t = \bar{y}_t$ with $\text{var}(\hat{\tau}_t) = \sigma^2/(n_t w_t)$, $t = 1, \dots, K$. We note that $\hat{\tau}_t$ is independent of $\hat{\tau}_s$, for $s \neq t$, and as before, if $n = \sum_{i=1}^K n_i$, the percent of subjects allocated to treatment t is $p_t = n_t/n$, with $\sum_{t=1}^K p_t = 1$.

We now further illustrate the flexibility of our approach by considering two design situations. In the first case, we seek an optimal design to minimize the average variance over all pairs $i - j$, $1 \leq i < j \leq K$, and in the second case, we wish to minimize the sum of variances of all pairwise comparisons that involve the placebo group. In the first case, we choose n_i 's to minimize

$$\frac{1}{\binom{K}{2}} \sum_{1 \leq i < j \leq K} \text{var}(\hat{\tau}_i - \hat{\tau}_j) = \frac{\sigma^2}{n \binom{K}{2}} \sum_{1 \leq i < j \leq K} \left(\frac{1}{p_i w_i} + \frac{1}{p_j w_j} \right).$$

If we let

$$F = \sum_{1 \leq i < j \leq K} \left(\frac{1}{p_i w_i} + \frac{1}{p_j w_j} \right) + \lambda \left(\sum_{i=1}^K p_i - 1 \right),$$

the Lagrange multiplier's method provides the optimal allocation proportions p_i , $i = 1, \dots, K$. There are $(K-1)$ terms in the first summation that contains p_i and they come from the variance of the contrasts $(\tau_i - \tau_j)$, $t = 1, \dots, i-1$, and the contrasts $(\tau_i - \tau_j)$, $t = i+1, \dots, K$. Equating the partial derivatives to zero gives

$$\frac{\partial F}{\partial p_i} = -\frac{K-1}{p_i^2 w_i} + \lambda = 0, i = 1, \dots, K.$$

It follows that

$$p_i = \sqrt{\frac{K-1}{\lambda w_i}}, i = 1, \dots, K$$

and using the fact that the p_i 's sum to unity, we arrive at the optimal proportions

$$p_i = \sqrt{\frac{K-1}{\lambda w_i}} = \frac{\sqrt{\frac{1}{w_i}}}{\sum_{t=1}^K \sqrt{\frac{1}{w_t}}}, i = 1, \dots, K.$$

We observe that the number of subjects assigned to each group by the A-optimal design is proportional to the standard deviation of the response from the group. This means that the A-optimal designs coincide with the designs obtained from the naive rule discussed in section 3.

In the second case, we wish to find an A-optimal design for comparing treatment groups with the control group. Assuming that the control group is the K^{th} group, an appropriate design criterion is to find n_i 's that minimize

$$\frac{1}{K-1} \sum_{i=1}^{K-1} \text{var}(\hat{\tau}_i - \hat{\tau}_K) = \frac{\sigma^2}{n(K-1)} \left[\left(\sum_{i=1}^{K-1} \frac{1}{p_i w_i} \right) + \frac{K-1}{p_K w_K} \right].$$

A similar argument shows that the optimal proportions are

$$p_i = \sqrt{\frac{1}{\lambda w_i}} = \frac{\sqrt{\frac{1}{w_i}}}{\sum_{t=1}^{K-1} \sqrt{\frac{1}{w_t}} + \sqrt{\frac{K-1}{w_t}}}, i=1, \dots, K-1 \quad p_K = \sqrt{\frac{K-1}{\lambda w_K}} = \frac{\sqrt{\frac{K-1}{w_K}}}{\sum_{t=1}^{K-1} \sqrt{\frac{1}{w_t}} + \sqrt{\frac{K-1}{w_t}}}.$$

When there are three or more treatment groups to compare, we modify the matrix A in Section 3 and apply the theory of optimal design to obtain a set of equations similar to (1) and (2). The optimality of the designs can be checked using the GET conditions for A-optimality [1].

Sometimes, it is appropriate to compare mean responses within treatment categories. For instance, the aim of the study may be to compare effectiveness of first line drugs for treating a certain disease and the secondary interest is in comparing effectiveness among second line drugs. Suppose there are K_1 first line drugs and K_2 second line drugs and within the same category, all comparisons are of equal interest and the variances are the same. Without loss of generality, let $w_i = 1, i = 1, \dots, K_1$ and $w_i = w, i = K_1 + 1, \dots, K, w > 0$ and $w \neq 1$. The proportions of patients to be assigned to treatments within the two categories are $p_i = p, i = 1, \dots, K_1$ for the first category and $p_i = q, i = K_1 + 1, \dots, K$, for the second category. We show in the appendix that conditions (i) and (ii) of GET imply the optimal value of p is

$$p = \frac{1}{2K_1(K-1)(1-w)} \left[(K_1-1)(1-w) - wK + \sqrt{\nabla} \right], \quad (3)$$

with $\nabla = [(K_1-1)(1-w) - wK]^2 + 4wK_1(K-1)(1-w)$ and $K_1p + (K-K_1)q = 1$.

In practice, not all comparisons may be of equal interest. Even when we have three treatments, the researcher may be more interested to estimate the mean difference between the first two treatments, and at the same time, like to have some information about how the third treatment compares with the first two. This suggests that the design should deliver a user-specified level of precision for estimates for the first comparison and also provides the best possible estimate for the second comparison, in terms of minimizing the variance of the estimate.

There are two methods for molding optimal designs to satisfy multiple objectives: constrained and compound optimal designs. These approaches are described in [13,14], for example. The latter type of optimal designs is found by minimizing a convex combination of convex functionals and so can be found using standard algorithms. Afterwards, the more easily interpretable constrained optimal design posed by the problem is found from among the class of compound optimal designs. We illustrate how to determine such an optimal design in the next section.

7. APPLICATIONS AND DISCUSSION

We provide two applications of our design strategies to find optimal designs in clinical trials. The first one uses real data from a trial at UCLA and the other demonstrates a potential application to design a cancer prevention trial.

Data from the Western Consortium of Rheumatologists study, an NIH-funded observational study at UCLA, shows some outcome variables from very early rheumatoid arthritis (RA) patients have variances that vary substantially from treatment protocol to protocol. There are 306 patients and seven treatment protocols in the study: (1) NSAIDS (non-steroidal anti-inflammatory drugs) alone, (2) NSAIDS + injectable gold, (3) NSAIDS + methotrexate, (4) NSAIDS + hydroxychloroquine (plaquinil), (5) NSAIDS + methotrexate +

hydroxychloroquine, (6) step-down bridge therapy and (7) NSAIDS + sulfasalazine (azulfidine).

C-reactive protein (CRP) and erythrocyte sedimentation rate (ESR) are common markers of inflammation and indicators of disease progression in RA patients. Either of these can be used as a component in the American College of Rheumatology (ACR) composite measure for disease improvement in RA patients. In numerical order of the above listed protocols, the sample sizes for the CRP are 24, 13, 139, 56, 23, 17 and 29. The corresponding sample standard deviations for CRP are 16, 5.29, 3.38, 3.43, 4.13, 1.61 and 3.31. A direct application of the Bartlett's test shows that the variances of the CRP are significantly different across protocols at the 0.001 significance level. If a trial is proposed to measure the effectiveness of the seven treatment protocols in controlling inflammation in RA patients and CRP is the primary outcome, our results in Section 3 show that the optimal allocation scheme is to allocate the following percentage of patients to the above protocols in the order prescribed: 16.55, 15.64, 14.35, 14.41, 15.04, 9.74 and 14.27. It is interesting to note that there are multiple solutions to the set of simultaneous equations defined by (1) and (2) in Section 3; however there is only a unique non-negative solution that is not larger than the prescribed upper bound " $1/(K-1)$ ". A similar situation arises for the outcome ESR. The explanation for the uniqueness is that the left hand side of equation (1) as a function of p_1 is continuous and strictly monotonic on the interval $(0, 1/(K-1))$, and it takes on opposite signs at the endpoints of this interval.

Our results have potential applications to design cancer control and prevention trials, and more generally to any individually randomized trial that has a continuous primary outcome whose variance is expected to vary across intervention groups. In cancer screening trials such as the one conducted by McPhee et al., a 3-arm intervention program was used to ascertain hepatitis B vaccination rate among 3–18 year old Vietnamese in USA [15]. The intervention arms were carried out in Dallas and Houston with a control group in Washington D. C. The Intervention program in Dallas involved a community-based effort to organize a community coalition to raise awareness about hepatitis B and in Houston, a Vietnamese language mass media campaign was used. One of their aims was to compare if knowledge of the disease is improved at the end of the trial. The variable knowledge is a continuous outcome and we may apply our techniques to construct an optimal design using two pairs of simple contrasts. The variable knowledge may or may not have different variability in the three areas; if they have similar variances, our optimal proportions will not be appreciably different and the equal allocation rule applies. In practice, different geographical areas may result in uncontrolled heterogeneity in the outcome variable. Our proposed designs provide a flexible option and will perform particularly well if preliminary data suggest there is variability across the groups. Additionally, our designs can incorporate the other objectives as in McPhee's study at the design stage and adjust the optimal allocation rule to accommodate the different level of interest in each comparison.

We do not have data from this trial and assume for illustrative purposes that (i) variances from the three groups are inversely proportional to $w_i = 1/i$, $i = 1, 2, 3$ and (ii) the primary goal is to compare mean difference between Houston and Washington D. C and the secondary goal is to compare all three mean differences. We set $K_1 = 2$, $K = 3$

$$A_1^T = (1 \ -1) \text{ and } A_2^T = \begin{bmatrix} 1 & -1 & 0 \\ 1 & 0 & -1 \end{bmatrix}.$$

Let λ be a weight used in the convex combination of the two convex criteria $\phi_1(\cdot)$ and $\phi_2(\cdot)$ chosen to represent the two objectives. Further, suppose the compound optimal design ξ_λ^*

assigns p_1 , p_2 and p_3 to the first, second and third group respectively. If $a_i = 1/(w_i p_i) = i/p_i$, $i = 1, 2, 3$, it is straightforward to verify that

$$|A_1^T [M^{-1}(\xi_\lambda^*)]_{11} A_1| = a_1 + a_2,$$

and

$$|A_2^T [M^{-1}(\xi_\lambda^*)] A_2| = a_1 a_2 a_3 \left(\sum_{i=1}^3 \frac{1}{a_i} \right).$$

The compound optimal design is found by minimizing the functional

$$\Phi(\xi_\lambda | \lambda) = \lambda \Phi_1(\xi_\lambda) + (1 - \lambda) \Phi_2(\xi_\lambda) = \lambda \ln(a_1 + a_2) + (1 - \lambda) \ln \left[a_1 a_2 a_3 \left(\sum_{i=1}^3 \frac{1}{a_i} \right) \right].$$

The numerical optimal design can now be found using software like Maple or Mathematica. If we denote the D_A -optimal design for each of the two objectives by ξ_λ^1 and ξ_λ^2 , the relative efficiencies of the compound optimal design ξ_λ^* for the two objectives are

$$E_1(\xi_\lambda^*) = \frac{|A_1^T M^{-1}(\xi_\lambda^1) A_1|}{|A_1^T [M^{-1}(\xi_\lambda^*)]_{11} A_1|}$$

and

$$E_2(\xi_\lambda^*) = \sqrt{\frac{|A_2^T [M^{-1}(\xi_\lambda^2)] A_2|}{|A_2^T [M^{-1}(\xi_\lambda^*)] A_2|}}.$$

Table VI shows the compound optimal designs for some values of λ . If $\lambda = 0.5$, the compound optimal design assigns the following proportions 0.302, 0.416 and 0.282 to the first, second and third group. The last two columns are the efficiencies of the compound optimal design for estimating the two objectives. In practice, the level of precision for the more important objective is pre-selected via the value of e_1 . For example, if the researcher wants the efficiency of the design for estimating the first comparison to be at least 0.95, the optimal sample proportions from Houston, Washington D. C and Dallas are 39.5%, 55.8% and 4.7% respectively. From the second last row in Table VI, this design provides 48.9% efficiency for estimating the second comparison.

The naive rule samples 24.1%, 34.1% and 41.8% from Houston, Washington D. C and Dallas respectively. The efficiencies of this design for estimating the first and second comparisons are $e_1^{na} \approx .582$ and $e_2^{na} \approx .996$. The equal allocation rule samples an equal number of subject from the three cities and have efficiencies $e_1^{eq} \approx .648$ and $e_2^{eq} \approx .982$ for estimating the two comparisons.

In conclusion, many group allocation rules in the literature lack theoretical justifications and are based on convenience or intuition. For example, the equal allocation rule is popular among researchers in part because it is simple to implement. When the method is theory based, the approach invariably restricts to a single objective. Current methods do not

adequately address design issues of variance heterogeneity in the outcome from different groups and multiple goals in the study. Our designs are based on optimal design theory, cost effective and provides clear advantages over current design strategies that ignores variance heterogeneity among various groups and the multiple objectives in the study. Our results here also give further support to findings reported in [16,17,18] that unbalanced designs are under-utilized in practice even though such designs can prove efficient.

Acknowledgments

The research of Wong and Zhu is partially supported from a NIH Grant R01GM072876. Additionally, the research of Wong is partially supported by a grant from The Scleroderma Foundation and NIH Grants R01GM072876, R01CA102486, P0109091 and P30CA16042-32. The research of Zhu is also partially supported by NIH grants M01RR010710, P30AG08051 and a DOE SciDAC grant.

APPENDIX: PROOFS OF RESULTS

Proof of results for the optimal rule

Under our setup, the information matrix $M(\xi)$ of a design ξ is proportional to a diagonal matrix with diagonal entries equal to $w_i p_i$, $i = 1, \dots, K$. Let A be the matrix defined in Section 3 and a direct calculation shows

$$A^T M^{-1}(\xi) A = \begin{bmatrix} \frac{1}{w_1 p_1} + \frac{1}{w_2 p_2} & \frac{1}{w_1 p_1} & \dots & \frac{1}{w_1 p_1} \\ \frac{1}{w_1 p_1} & \frac{1}{w_1 p_1} + \frac{1}{w_3 p_3} & \dots & \frac{1}{w_1 p_1} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{1}{w_1 p_1} & \frac{1}{w_1 p_1} & \dots & \frac{1}{w_1 p_1} + \frac{1}{w_K p_K} \end{bmatrix}$$

If $\Delta = \sum_{i=1}^K w_i p_i$, we have

$$B = A \left\{ A^T M^{-1}(\xi) A \right\}^{-1} A^T = \begin{bmatrix} w_1 p_1 - \frac{(w_1 p_1)^2}{\Delta} & -\frac{(w_1 p_1)(w_2 p_2)}{\Delta} & \dots & -\frac{(w_1 p_1)(w_K p_K)}{\Delta} \\ -\frac{(w_1 p_1)(w_2 p_2)}{\Delta} & w_2 p_2 - \frac{(w_2 p_2)^2}{\Delta} & \dots & -\frac{(w_2 p_2)(w_K p_K)}{\Delta} \\ \vdots & \vdots & \ddots & \vdots \\ -\frac{(w_1 p_1)(w_K p_K)}{\Delta} & -\frac{(w_2 p_2)(w_K p_K)}{\Delta} & \dots & w_K p_K - \frac{(w_K p_K)^2}{\Delta} \end{bmatrix}.$$

and it follows that

$$d_A(x_i, \xi) = w_i \cdot f^T(x_i) M^{-1}(\xi) B M^{-1}(\xi) f(x_i) = w_i \cdot \left(\frac{1}{w_i p_i} \right)^2 \left[w_i p_i - \frac{(w_i p_i)^2}{\Delta} \right] = \frac{1}{p_i} - \frac{w_i}{\Delta}, i=1, \dots, K.$$

The optimal design ξ^* can be obtained by solving $d_A(x_i, \xi) = K-1$, $i=1, \dots, K$, which after simplifications reduces to equations (1) and (2). The upper bound on p_i is obtained from the same equation where after rewriting, we have $1 - (K-1)p_i = w_i p_i / \sum_{i=1}^K w_i p_i \geq 0$.

Proof of results in the two-category scenario

We substitute $p_1 = p$, $w_i = 1$, $i=1, \dots, K_1$ and $w_i = w$, $i=K_1+1, \dots, K$ into equations (1) and (2) and solve for p . This results in a quadratic equation

$$K_1(K-1)\left(\frac{1}{w}-1\right)p^2 + [K - (K_1-1)\left(\frac{1}{w}-1\right)]p - 1 = 0.$$

When $w = 1$; the solution is $p = 1/K$ as expected. When $w \neq 1$, we have two solutions

$$p_+ = \frac{1}{2K_1(K-1)(1-w)} \left[(K_1-1)(1-w) - wK + \sqrt{\Delta} \right],$$

and

$$p_- = \frac{1}{2K_1(K-1)(1-w)} \left[(K_1-1)(1-w) - wK - \sqrt{\Delta} \right],$$

where $\Delta = [(K_1-1)(1-w) - wK]^2 + 4wK_1(K-1)(1-w)$. The following argument shows p_+ is the desired solution.

1. When $0 < w < (K_1-1)/(K+K_1-1)$, $2K_1(K_1-1)(1-w) > 0$, and $0 < (K_1-1)(1-w) - wK < \sqrt{\Delta}$, hence $p_- < 0$; so choose p_+ ;
2. When $(K_1-1)/(K+K_1-1) \leq w < 1$, $2K_1(K-1)(1-w) > 0$ and $(K_1-1)(1-w) - wK \leq 0$, hence $p_- < 0$; so choose p_+ ;
3. When $w > 1$, $2K_1(K-1)(1-w) < 0$ and $(K_1-1)(1-w) - wK < 0$. Both p_+ and p_- are positive, but it is easily seen that $p_+ < p_-$ and $p_+ + p_- > 2/(K-1)$, hence $p_- > 1/(K-1)$, exceeds the upper bound, see results for the optimal rule. Therefore the desired solution can only be p_+ .

Combining the three scenarios above, we found that the desired solution is $p = p_+$.

REFERENCES

1. Atkinson, AC.; Donev, AN. Optimum Experimental Designs. Oxford Science Publications; 1992.
2. Kiefer, J. Jack Carl Kiefer Collected Papers III: Design of Experiments. New York: Springer Verlag; 1980.
3. Atkinson AC. Optimum biased coin designs for sequential trials with prognostic factors. *Biometrika*. 1982; 69:61–67.
4. Zhu W, Wong WK. Optimum treatment allocation in for dual-objective clinical trials with binary outcomes. *Communications in Statistics-Theory and Methods*. 2000; 29:957–974.
5. Dette H, Wong WK, Zhu W. On the equivalence of optimality design criteria for the placebo treatment problem. *Statistics and Probability Letters*. 2005; 74:337–346.
6. Marschner IC. Optimal design of clinical trials comparing several treatments with a control. *Pharmaceutical Statistics*. 2007; 6:23–34.
7. Atkinson AC. Optimum biased-coin designs for sequential treatment allocation with covariate information. *Statistics in Medicine*. 1999; 18:1741–1752. [PubMed: 10407244]
8. Silvey, SD. Optimal design. London, New York: Chapman and Hall; 1980.
9. Mendenhall, W.; Scheaffer, RL.; Wackerly, DD. Mathematical statistics with applications. 3rd. Edition. Boston: Duxbury Press; 1986.
10. Dette H. On robust and efficient design for risk estimation in epidemiological studies. *Scandinavian Journal of Statistics*. 2004; 31:319–331.

11. Berger, MPF.; Ouwens, MJJN.; Tan, FES. Robust designs for longitudinal mixed effects models. In: Yanai, H., et al., editors. *Advances in Psychometric Research*. Springer Verlag; 2002. p. 339-346.
12. Berger MPF, Tan FES. Robust designs for linear mixed effects models. *Journal of the Royal Statistical Society, series C*. 2004; 53:569–581.
13. Cook RD, Wong WK. On the equivalence of constrained and compound optimal designs. *Journal of the American Statistical Association*. 1994; 89:687–692.
14. Imhof L, Wong WK. A graphical method for finding maximin designs. *Biometrics*. 2000; 56:113–117. [PubMed: 10783784]
15. McPhee SJ, Nguyen T, Mock GL, Wong C, Lam T, Nguyen W, Nguyen S, Ha MQH, Do ST, Buu C. Successful promotion of hepatitis B vaccinations among Vietnamese-American children ages 3–18: results of a controlled trial. *Pediatrics*. 2003; 111:1278–1288. [PubMed: 12777542]
16. Dumville JC, Hahn S, Miles JNV, Torgerson DJ. The use of unequal randomisation ratios in clinical trials: a review. *Contemporary Clinical Trials*. 2006; 27:1–12. [PubMed: 16236557]
17. Hu F, Rosenberger WF, Zhang LX. Asymptotically best response-adaptive randomization procedures. *Journal of Statistical Planning and Inference*. 2006; 136:1911–1922.
18. Tymofyeyev Y, Rosenberger WF, Hu F. Implementing optimal allocation in sequential binary response experiments. *Journal of the American Statistical Association*. 2007; 102:224–234.

Table I

Comparison of the three rules for scenario 2.

Weights (w_1, w_2, w_3)	D_A -optimal ($p, p, 1-2p$)	Naive rule ($p_1^{\text{na}}, p_2^{\text{na}}, p_3^{\text{na}}$)	e^{na}	e^{eq}
$(1, 1, \frac{1}{2})$	(0.309, 0.309, 0.382)	(0.293, 0.293, 0.414)	0.997	0.993
$(1, 1, \frac{1}{4})$	(0.289, 0.289, 0.422)	(0.250, 0.250, 0.500)	0.985	0.979
$(1, 1, \frac{1}{8})$	(0.273, 0.273, 0.453)	(0.207, 0.207, 0.586)	0.958	0.965
$(1, 1, \frac{1}{16})$	(0.263, 0.263, 0.474)	(0.167, 0.167, 0.666)	0.915	0.955
$(1, 1, \frac{1}{32})$	(0.257, 0.257, 0.486)	(0.131, 0.131, 0.738)	0.855	0.950
$(1, 1, \frac{1}{64})$	(0.254, 0.254, 0.492)	(0.100, 0.100, 0.800)	0.782	0.946

Table II

Comparison of the three rules for scenario 3.

Weights (w_1, w_2, w_3)	D_A -optimal ($p, p, 1-2p$)	Naive rule ($p_1^{\text{na}}, p_2^{\text{na}}, p_3^{\text{na}}$)	e^{na}	e^{eq}
(1, 1, 2)	(0.360, 0.360, 0.280)	(0.369, 0.369, 0.262)	0.998	0.990
(1, 1, 4)	(0.386, 0.386, 0.228)	(0.400, 0.400, 0.200)	0.996	0.958
(1, 1, 8)	(0.410, 0.410, 0.180)	(0.425, 0.425, 0.150)	0.994	0.911
(1, 1, 16)	(0.430, 0.430, 0.140)	(0.444, 0.444, 0.112)	0.993	0.860
(1, 1, 32)	(0.447, 0.447, 0.106)	(0.459, 0.459, 0.082)	0.993	0.813
(1, 1, 64)	(0.461, 0.461, 0.078)	(0.471, 0.471, 0.058)	0.994	0.775

Table III

Comparison of the three rules for scenario 4.

Weights (w_1, w_2, w_3)	D_A -optimal (p_1, p_2, p_3)	Naive rule ($p_1^{\text{na}}, p_2^{\text{na}}, p_3^{\text{na}}$)	e^{na}	e^{eq}
$(1, \frac{1}{4}, \frac{1}{16})$	(0.190, 0.356, 0.454)	(0.143, 0.286, 0.571)	0.966	0.921
$(1, \frac{1}{8}, \frac{1}{16})$	(0.162, 0.396, 0.442)	(0.128, 0.361, 0.511)	0.985	0.888
$(1, \frac{1}{8}, \frac{1}{32})$	(0.148, 0.386, 0.466)	(0.106, 0.298, 0.596)	0.959	0.872
$(1, \frac{1}{16}, \frac{1}{32})$	(0.124, 0.420, 0.456)	(0.094, 0.375, 0.531)	0.984	0.839
$(1, \frac{1}{2}, \frac{1}{16})$	(0.225, 0.310, 0.465)	(0.156, 0.220, 0.624)	0.941	0.947
$(1, \frac{1}{2}, \frac{1}{32})$	(0.217, 0.303, 0.480)	(0.124, 0.175, 0.701)	0.890	0.939
$(1, \frac{1}{4}, \frac{1}{32})$	(0.180, 0.346, 0.474)	(0.116, 0.231, 0.653)	0.926	0.909

Table IV

Allocation rules for four groups with 1, 2, 3 or 4 unequal variances.

Weights	Optimal design	Naive rule	e^{na}	e^{eq}
(1, 1, 1, 4)	(0.274, 0.274, 0.274, 0.178)	(0.286, 0.286, 0.286, 0.143)	0.992	0.978
(1, 1, 1, 8)	(0.286, 0.286, 0.286, 0.142)	(0.298, 0.298, 0.298, 0.105)	0.988	0.948
(1, 1, 1, 16)	(0.296, 0.296, 0.296, 0.112)	(0.308, 0.308, 0.308, 0.077)	0.986	0.912
$(1, 1, 1, \frac{1}{2})$	(0.240, 0.240, 0.240, 0.279)	(0.227, 0.227, 0.227, 0.320)	0.995	0.997
$(1, 1, 1, \frac{1}{4})$	(0.233, 0.233, 0.233, 0.301)	(0.200, 0.200, 0.200, 0.400)	0.976	0.992
$(1, 1, 1, \frac{1}{8})$	(0.228, 0.228, 0.228, 0.316)	(0.172, 0.172, 0.172, 0.485)	0.937	0.988
$(1, 1, 1, \frac{1}{16})$	(0.225, 0.225, 0.225, 0.325)	(0.143, 0.143, 0.143, 0.571)	0.877	0.986
$(1, 1, \frac{1}{2}, \frac{1}{2})$	(0.229, 0.229, 0.271, 0.271)	(0.207, 0.207, 0.293, 0.293)	0.995	0.995
$(1, 1, \frac{1}{4}, \frac{1}{4})$	(0.210, 0.210, 0.290, 0.290)	(0.369, 0.369, 0.262, 0.262)	0.980	0.984
$(1, 1, \frac{1}{8}, \frac{1}{8})$	(0.194, 0.194, 0.306, 0.306)	(0.400, 0.400, 0.200, 0.200)	0.954	0.971
$(1, 1, \frac{1}{16}, \frac{1}{16})$	(0.183, 0.183, 0.317, 0.317)	(0.425, 0.425, 0.150, 0.150)	0.917	0.961
$(1, 1, \frac{1}{32}, \frac{1}{32})$	(0.176, 0.176, 0.324, 0.324)	(0.444, 0.444, 0.112, 0.112)	0.870	0.954
(1, 1, 2, 16)	(0.301, 0.301, 0.275, 0.123)	(0.338, 0.338, 0.239, 0.085)	0.979	0.927
(1, 1, 4, 16)	(0.307, 0.307, 0.247, 0.139)	(0.364, 0.364, 0.182, 0.091)	0.961	0.943
(1, 1, 4, 32)	(0.313, 0.313, 0.265, 0.109)	(0.374, 0.374, 0.187, 0.066)	0.956	0.909
(1, 1, 8, 32)	(0.317, 0.317, 0.238, 0.128)	(0.395, 0.395, 0.140, 0.070)	0.927	0.931
$(1, 1, \frac{1}{4}, \frac{1}{16})$	(0.198, 0.198, 0.284, 0.320)	(0.125, 0.125, 0.250, 0.500)	0.919	0.974
$(1, 1, \frac{1}{8}, \frac{1}{32})$	(0.186, 0.186, 0.303, 0.325)	(0.095, 0.095, 0.270, 0.540)	0.878	0.963
$(1, 1, \frac{1}{16}, \frac{1}{32})$	(0.179, 0.179, 0.317, 0.325)	(0.086, 0.086, 0.343, 0.485)	0.884	0.957
$(1, \frac{1}{2}, \frac{1}{4}, \frac{1}{8})$	(0.184, 0.237, 0.277, 0.303)	(0.138, 0.195, 0.276, 0.391)	0.974	0.978

Weights	Optimal design	Naive rule	e^{na}	e^{eq}
$\frac{1}{1}, \frac{1}{4}, \frac{1}{12}$ (1, 4, 8, 12)	(0.151, 0.256, 0.290, 0.303)	(0.108, 0.215, 0.304, 0.373)	0.975	0.955
$\frac{1}{1}, \frac{1}{8}, \frac{1}{32}$ (1, 8, 24, 32)	(0.111, 0.267, 0.308, 0.314)	(0.070, 0.197, 0.341, 0.393)	0.960	0.913
$\frac{1}{1}, \frac{1}{16}, \frac{1}{32}$ (1, 16, 24, 32)	(0.098, 0.290, 0.303, 0.310)	(0.064, 0.257, 0.315, 0.364)	0.979	0.895
$\frac{1}{1}, \frac{1}{2}, \frac{1}{16}$ (1, 2, 4, 16)	(0.178, 0.232, 0.274, 0.316)	(0.119, 0.168, 0.238, 0.475)	0.938	0.973
$\frac{1}{1}, \frac{1}{2}, \frac{1}{16}$ (1, 2, 16, 32)	(0.155, 0.212, 0.311, 0.322)	(0.083, 0.117, 0.331, 0.469)	0.910	0.953
$\frac{1}{1}, \frac{1}{2}, \frac{1}{8}$ (1, 2, 8, 32)	(0.163, 0.219, 0.295, 0.323)	(0.092, 0.130, 0.260, 0.519)	0.902	0.960

Table VEfficiencies of D_A -optimal designs under variances mis-specification.

Real weights (w_1, w_2, w_3)	Assumed weights (w_1, w_2, w_3)	D_A -efficiency
$(1, \frac{1}{2}, \frac{1}{16})$	$(1, 1, \frac{1}{8})$	0.986
$(1, \frac{1}{2}, \frac{1}{16})$	$(1, 1, \frac{1}{4})$	0.983
$(1, \frac{1}{2}, \frac{1}{16})$	$(1, 1, \frac{1}{2})$	0.972
$(1, \frac{1}{2}, \frac{1}{16})$	$(1, 1, 1)$	0.946
$(1, \frac{1}{2}, \frac{1}{16})$	$(1, 1, 2)$	0.906
$(1, \frac{1}{2}, \frac{1}{16})$	$(1, 1, 4)$	0.850
$(1, \frac{1}{2}, \frac{1}{16})$	$(1, 2, 16)$	0.803

Table VI

Constrained and compound optimal designs and their efficiencies for the dual objective design problem.

	p_1	p_2	p_3	e_1	e_2
0	0.266	0.347	0.387	0.612	1
0.25	0.280	0.376	0.344	0.656	0.995
0.50	0.302	0.416	0.282	0.718	0.966
0.78	0.347	0.487	0.166	0.833	0.833
0.85	0.364	0.513	0.123	0.877	0.743
0.95	0.395	0.558	0.047	0.953	0.489
1	0.414	0.586	0	1	0