Biometrics 73, 916–926
September 2017

DOI: 10.1111/biom.12659

Maximin Optimal Designs for Cluster Randomized Trials

Sheng Wu,* Weng Kee Wong, and Catherine M. Crespi

Department of Biostatistics, UCLA Fielding School of Public Health, University of California, Los Angeles, California 90095-1772, U.S.A.

**email:* shengwu@ucla.edu

Summary. We consider design issues for cluster randomized trials (CRTs) with a binary outcome where both unit costs and intraclass correlation coefficients (ICCs) in the two arms may be unequal. We first propose a design that maximizes cost efficiency (CE), defined as the ratio of the precision of the efficacy measure to the study cost. Because such designs can be highly sensitive to the unknown ICCs and the anticipated success rates in the two arms, a local strategy based on a single set of best guesses for the ICCs and success rates can be risky. To mitigate this issue, we propose a maximin optimal design that permits ranges of values to be specified for the success rate and the ICC in each arm. We derive maximin optimal designs for three common measures of the efficacy of the intervention, risk difference, relative risk and odds ratio, and study their properties. Using a real cancer control and prevention trial example, we ascertain the efficiency of the widely used balanced design relative to the maximin optimal design and show that the former can be quite inefficient and less robust to mis-specifications of the ICCs and the success rates in the two arms.

Key words: Balanced design; Binary outcome; Intraclass correlation coefficient; Relative cost efficiency; Robust design; Sampling ratio.

1. Introduction

Cluster randomized trials (CRTs) are increasingly used in many fields including public health, education, and clinical research (Donner and Klar, 2000; Hayes and Moulton, 2009). CRTs are experiments in which clusters of individuals rather than independent individuals are randomly allocated to intervention groups. All individuals in a given cluster receive the same treatment. Clusters can be churches, villages, medical practices, families, or schools. A key feature of CRTs is that outcomes of individuals within a cluster are correlated. The intraclass correlation coefficient (ICC), usually denoted by ρ, provides a quantitative measure of within-cluster correlation. The ICC is variously defined as the Pearson correlation between two members in the same cluster or the proportion of the total variance in the outcome attributable to the variance between clusters. Since the correlation increases the sampling error of estimating the intervention effect (Donner, Birkett, and Buck, 1981), CRTs are less efficient than individual randomized trials (IRTs). However, there are many reasons to use CRTs, including administrative convenience, ethical considerations, to avoid treatment group contamination and because the intervention is naturally applied at the cluster level.

All else being equal, investigators prefer to expend minimal resources to obtain the most accurate estimate of an intervention effect. This is even more pertinent when designing CRTs because CRTs can be much less efficient than IRTs (see, e.g., Donner and Klar, 2000). However, because of the correlated data structure, design issues for CRTs are more complicated than for IRTs (Moerbeek and Teerenstra, 2016). In practice, investigators usually use a two-arm CRT and assign the same number of clusters to each arm (Hayes and Moulton, 2009). Following classic analysis of variance terminology (for e.g., Milliken and Johnson, 1984), we call such a design a balanced design. Previous research on IRTs has shown that a balanced design may not be the most efficient, particularly when costs are unequal in the two arms; discussions can be found in Yanagawa and Bolt (1977), Meydrich (1978), Lubin (1980), Brittain and Schlesselman (1982), Morgenstern and Winn (1983), and Gail et al. (1996). Several authors, including Raudenbush (1997), Moerbeek, Breukelen and Berger (2000), Raudenbush and Liu (2000), Breukelen and Candel (2012), and Moerbeek and Teerenstra (2016), have discussed optimal design issues for CRTs that included cost considerations in their optimality criterion. However, they have focused mainly on finding optimal sample size per cluster rather than optimal allocation of clusters to the two or more conditions. Their designs assume an equal number of clusters in the two arms. In addition, they assume the outcomes are continuous and the ICCs are the same in the two arms.

The expected success rates in the different conditions are important parameters for any IRT or CRT design. Dette (2004) noted that almost all optimal designs for IRTs are locally optimal in that they depend on the unknown success rates. Consequently, such designs may not be robust when success rates are mis-specified. He proposed a maximin method to construct designs that are robust with respect to the unknown parameters. His idea was to find a maximin optimal design that maximizes the minimum efficiency over a plausible region of nominal possible values of the parameters. He provided some theoretical justifications but had no real application.

Our aim in this article is to develop a flexible maximin approach for designing a two-arm CRT with binary outcomes. We assume the total number of clusters is fixed in advance and the objective is to determine the optimal proportion of clusters to assign to each arm, considering costs. Often, CRTs involve a fixed predetermined number of clusters, due to constraints on recruitment rate or the number of available clusters, or financial constraints. Such a maximin optimal design offers some global protection against the worst case scenario when the nominal values of the parameters for the design problem are very incorrect. We allow both costs and ICCs to vary between the two arms, and we develop the approach for the three most common treatment effect measures for binary outcomes, risk difference (RD), relative risk (RR), and odds ratio (OR). Cluster sizes are assumed equal. Using a cancer control and prevention trial, we illustrate that the balanced design that assigns an equal number of clusters to each arm can have low statistical and cost efficiencies.

The organization of this article is as follows: In Section 2, we introduce the common correlation model and define cost efficiency (CE). We derive the optimal allocations for estimating RD, RR, and OR by maximizing CE. We then define relative cost efficiency (RCE) and show that the RCEs of balanced designs compared to the optimal allocation can be low in many situations. Since the optimal allocation can be highly sensitive to the unknown ICCs and the anticipated success rates, a locally optimal design based on single best guesses for the ICCs and success rates can be risky. In Section 3, we propose a maximin optimal design that permits a range of values to be specified for the success rate and the ICC in each arm. In Section 4, we provide guidance on applying the methods and illustrate using a real CRT, and show that the maximin optimal design is generally more efficient (i.e., has a larger RCE) than the balanced design and is robust to mis-specifications of the ICCs and the success rates in the two arms. Section 5 provides a discussion. In the Web Appendix, we provide a proof of our main result for the maximin approach, sensitivity analyzes, and R code to implement the proposed maximin optimal designs for user-specified settings.

2. Optimal Allocation

Our two-arm CRTs with binary outcomes are based on the common correlation model; see, for example, Ridout, Demetrio, and Firth (1999) and Eldridge, Ukoumunne, and Carlin (2009). Let X_{hij} denote the response of the jth individual in the ith cluster in the hth treatment arm. Let $X_{hij}=1$ when the outcome of interest is present (success) and $X_{hij}=0$ otherwise (failure). We assume that the success rate $Pr(X_{hij}=1)$ for all individuals in all clusters in the hth treatment arm is the same and equal to π_h, $h\in\{1,2\}$ and all cluster sizes are equal to m. The responses of individuals from different clusters are assumed to be independent, and within each cluster, the correlation of responses between any pair of individuals is ρ_{hi}, the ICC. We further assume that (i) the ICCs for all clusters in the hth treatment arm are the same, so the subscript i in ρ_{hi} can be removed, (ii) the total number of clusters in the trial is predetermined and equal to k; k_1, k_2 are the numbers of clusters in arms 1 and 2, respectively, such that $k=k_1+k_2$, and (iii) ρ_1 is not necessarily equal to ρ_2. The last assumption is more flexible and also more realistic in some intervention trials; see for example, Crespi, Wong, and Mishra (2009), Crespi, Wong, and Wu (2011), and Wu, Crespi, and Wong (2012).

We consider three commonly used treatment effect measures, $RD=\pi_1-\pi_2$, $RR=\pi_1/\pi_2$, and $OR=\frac{\pi_1/(1-\pi_1)}{\pi_2/(1-\pi_2)}$. For a given measure, our goal is to determine the optimal proportion of clusters to allocate to arm 1, $w=k_1/k$, in order to minimize the asymptotic variance of the relevant estimator, $\hat{RD}=\hat\pi_1-\hat\pi_2$, $\hat{RR}=\hat\pi_1/\hat\pi_2$, or $\hat{OR}=\frac{\hat\pi_1/(1-\hat\pi_1)}{\hat\pi_2/(1-\hat\pi_2)}$. The allocation scheme that minimizes this variance is called the optimal allocation. The variances can be derived as follows. By the central limit theorem, the maximum-likelihood estimates of $(\hat\pi_1,\hat\pi_2)$ for the success rates are approximately normal with

$$\sqrt{k}\begin{pmatrix}\hat\pi_1-\pi_1\\ \hat\pi_2-\pi_2\end{pmatrix}\xrightarrow{D} N\left\{\begin{pmatrix}0\\0\end{pmatrix},\begin{pmatrix}\dfrac{\pi_1(1-\pi_1)d_1}{wm} & 0\\ 0 & \dfrac{\pi_2(1-\pi_2)d_2}{(1-w)m}\end{pmatrix}\right\},$$

where $d_h=1+(m-1)\rho_h$ is the design effect for arm $h\in\{1,2\}$. The asymptotic variance of $\hat{RD}$ is

$$\Psi_{RD}^{-1}=\pi_1(1-\pi_1)\frac{d_1}{m}\left[\frac{1}{w}+\frac{\pi_2(1-\pi_2)d_2}{\pi_1(1-\pi_1)d_1(1-w)}\right],$$

and applying the delta method, we obtain asymptotic variance estimates for $\hat{RR}$ and $\hat{OR}$ as:

$$\Psi_{RR}^{-1}\approx\frac{\pi_1(1-\pi_1)}{\pi_2^2}\frac{d_1}{m}\left[\frac{1}{w}+\frac{\pi_1(1-\pi_2)d_2}{\pi_2(1-\pi_1)d_1(1-w)}\right]$$

and

$$\Psi_{OR}^{-1}\approx\frac{\pi_1(1-\pi_2)^2}{(1-\pi_1)^3\pi_2^2}\frac{d_1}{m}\left[\frac{1}{w}+\frac{\pi_1(1-\pi_1)d_2}{\pi_2(1-\pi_2)d_1(1-w)}\right].$$

Next, we consider study costs. In CRTs, there can be costs per individual and costs per cluster, and these could vary by arm. Let the cost per individual be c_h and the cost per cluster be e_h in arm h. The total cost function when each cluster has size m is

$$\begin{aligned}&k_1(mc_1+e_1)+k_2(mc_2+e_2)\\&\quad=k[w(mc_1+e_1)+(1-w)(mc_2+e_2)].\end{aligned}$$

Following Dette (2004), we define cost efficiency (CE) as the ratio of the precision of the treatment effect measure to the total study cost. This is a natural way to combine statistical and cost considerations. For each outcome measure $x\in\{RD,RR,OR\}$, the goal then is to determine the optimal proportion of clusters to assign to arm 1, denoted w_x^*, by maximizing the CE for measure x, given by

$$CE_x=\frac{\Psi_x}{k_1(mc_1+e_1)+k_2(mc_2+e_2)}.$$

To find this design, CE_x is optimized with respect to w by setting its first derivative equal to zero and solving for w. Defining the ratio of total per-cluster costs in the two arms

Table 1

RCE of a balanced design ($w = 0.5$) compared to the optimal design for estimating RD with fixed number of clusters under different combinations of π_1 and π_2 when $\rho_1 = 0.05$, $\rho_2 = 0.1$, $m = 20$, and $\gamma = 5(2)$

π_1/π_2	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
0.1	0.80 (.93)	0.68 (.85)	0.63 (.80)	0.60 (.78)	0.59 (.77)	0.60 (.78)	0.63 (.80)	0.68 (.85)	0.80 (.93)
0.2	0.90 (.98)	0.80 (.93)	0.75 (.89)	0.72 (.87)	0.71 (.87)	0.72 (.87)	0.75 (.89)	0.80 (.93)	0.90 (.98)
0.3	0.94 (1.00)	0.85 (.96)	0.80 (.93)	0.77 (.91)	0.77 (.91)	0.77 (.91)	0.80 (.93)	0.85 (.96)	0.94 (1.00)
0.4	0.95 (1.00)	0.87 (.97)	0.83 (.95)	0.80 (.93)	0.79 (.92)	0.80 (.93)	0.83 (.95)	0.87 (.97)	0.95 (1.00)
0.5	0.96 (1.00)	0.88 (.98)	0.83 (.95)	0.81 (.94)	0.80 (.93)	0.81 (.94)	0.83 (.95)	0.88 (.98)	0.96 (1.00)
0.6	0.95 (1.00)	0.87 (.97)	0.83 (.95)	0.80 (.93)	0.79 (.92)	0.80 (.93)	0.83 (.95)	0.87 (.97)	0.95 (1.00)
0.7	0.94 (1.00)	0.85 (.96)	0.80 (.93)	0.77 (.91)	0.77 (.91)	0.77 (.91)	0.80 (.93)	0.85 (.96)	0.94 (1.00)
0.8	0.90 (.98)	0.80 (.93)	0.75 (.89)	0.72 (.87)	0.71 (.87)	0.72 (.87)	0.75 (.89)	0.80 (.93)	0.90 (.98)
0.9	0.80 (.93)	0.68 (.85)	0.63 (.80)	0.60 (.78)	0.59 (.77)	0.60 (.78)	0.63 (.80)	0.68 (.85)	0.80 (.93)

as $\gamma = \frac{mc_1+e_1}{mc_2+e_2}$, it follows directly that the optimal allocation w_x^* that maximizes CE for each measure is

$$w_{RD}^* = \frac{\sqrt{\pi_1(1-\pi_1)d_1}}{\sqrt{\pi_1(1-\pi_1)d_1} + \sqrt{\gamma\pi_2(1-\pi_2)d_2}},$$

$$w_{RR}^* = \frac{\sqrt{\pi_2(1-\pi_1)d_1}}{\sqrt{\pi_2(1-\pi_1)d_1} + \sqrt{\gamma\pi_1(1-\pi_2)d_2}},$$

and

$$w_{OR}^* = \frac{\sqrt{\pi_2(1-\pi_2)d_1}}{\sqrt{\pi_2(1-\pi_2)d_1} + \sqrt{\gamma\pi_1(1-\pi_1)d_2}}.$$

We note that if $\rho_1 = \rho_2$, we have $d_1 = d_2$ and the optimal allocations for all three measures reduce to those reported in Dette (2004) for IRTs.

For a vector of fixed design parameters $\theta^T = (\pi_1, \pi_2, \rho_1, \rho_2)$, the design with a larger CE is more desirable, all else being equal. To compare different designs, we use relative cost efficiency (RCE), defined as the cost efficiency of a design with allocation w relative to the cost efficiency of the optimal design, that is, $RCE_x(w) = \frac{CE_x(w)}{CE_x(w_x^*)}$. The maximal value of RCE is 1, which is reached when w is the optimal allocation w_x^*. For a balanced design, $w = 0.5$. If $RCE_x(0.5)$ is close to 1, the balanced design performs about as well as the optimal design.

For different measures x, RCE of a balanced design compared to the optimal design can be quite different. Tables 1, 2, and 3 show $RCE_x(0.5)$ for estimating RD, RR, and OR, respectively, for different combinations of π_1 and π_2 when the total number of clusters in the trial is fixed. The value of the cost ratio, $\gamma = 5$, is motivated by one of our cancer control and prevention trials described more fully later. For illustration purposes, we also consider $\gamma = 2$ to ascertain whether $RCE_x(0.5)$ is sensitive to the cost ratio value. We focus here on the performance of the balanced design because it is widely used in practice. For space consideration, we only show the case when $\rho_1 = 0.05$, $\rho_2 = 0.1$, and $m = 20$, but interested readers can compute the RCE for any design of interest using the R code in Web Appendix A.4.

Table 1 shows $RCE_x(0.5)$ values when the treatment effect measure is RD. The RCEs are symmetrical about $\pi_1 = 0.5$ and about $\pi_2 = 0.5$ because $\pi_h(1-\pi_h)$, which is symmetrical about 0.5, appears in the formula. The RCEs range between 0.59 and 0.96 for $\gamma = 5$ and between 0.77 and 1.00 for $\gamma = 2$. In most scenarios, $RCE_x(0.5)$ is larger than 0.7 for $\gamma = 5$ and larger than 0.8 for $\gamma = 2$. The smallest value occurs when $\pi_1 = 0.1$ or 0.9, $\pi_2 = 0.5$ and $\gamma = 5$. Hence the balanced design performs satisfactorily in some cases but can be inefficient when costs or success rates are very different between arms.

Table 2 shows $RCE_x(0.5)$ values when the treatment effect measure is RR. Here, the RCEs are symmetrical about the diagonal line $\pi_1 = 1 - \pi_2$, which is also a direct consequence of the formula. The RCEs range between 0.24 and 1.00 for $\gamma = 5$ and between 0.42 and 1.00 for $\gamma = 2$. RCE values are smaller than 0.8 in many scenarios. This suggests that a balanced design often will not perform well for estimating RR. The smallest RCE of 0.24 occurs when $\pi_1 = 0.9$ and $\pi_2 = 0.1$. Although this magnitude of difference in success rates is unlikely to occur in practice, it shows that in extreme cases when the intervention arm is much more successful compared with the control arm, the balanced design can perform substantially worse for estimating RR than for estimating RD. This reinforces the recommendation that the design should be chosen appropriately for the outcome measure.

Table 3 shows $RCE_x(0.5)$ values for estimating the OR. Similar to RD, the RCEs are symmetrical about $\pi_1 = 0.5$ and about $\pi_2 = 0.5$. However, the peaks and trends are different because w_{RD}^* contains $\pi_1(1-\pi_1)$ in the numerator whereas

Table 2

RCE of a balanced design ($w = 0.5$) compared to the optimal design for estimating RR with fixed number of clusters under different combinations of π_1 and π_2 when $\rho_1 = 0.05$, $\rho_2 = 0.1$, $m = 20$, and $\gamma = 5(2)$

π_1/π_2	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
0.1	0.80 (.93)	0.93 (1.00)	0.98 (1.00)	1.00 (.98)	1.00 (.95)	0.99 (.92)	0.97 (.88)	0.95 (.84)	0.92 (.79)
0.2	0.63 (.81)	0.80 (.93)	0.90 (.98)	0.95 (1.00)	0.98 (1.00)	1.00 (.98)	1.00 (.94)	0.98 (.90)	0.95 (.84)
0.3	0.53 (.72)	0.69 (.85)	0.80 (.93)	0.88 (.98)	0.94 (1.00)	0.98 (1.00)	1.00 (.98)	1.00 (.94)	0.97 (.88)
0.4	0.46 (.65)	0.60 (.78)	0.71 (.87)	0.80 (.93)	0.87 (.97)	0.93 (1.00)	0.98 (1.00)	1.00 (.98)	0.99 (.92)
0.5	0.40 (.60)	0.52 (.71)	0.63 (.80)	0.72 (.87)	0.80 (.93)	0.87 (.97)	0.94 (1.00)	0.98 (1.00)	1.00 (.95)
0.6	0.36 (.55)	0.46 (.65)	0.55 (.73)	0.63 (.81)	0.72 (.87)	0.80 (.93)	0.88 (.98)	0.95 (1.00)	1.00 (.98)
0.7	0.32 (.51)	0.40 (.59)	0.47 (.66)	0.55 (.73)	0.63 (.80)	0.71 (.87)	0.80 (.93)	0.90 (.98)	0.98 (1.00)
0.8	0.28 (.47)	0.34 (.53)	0.40 (.59)	0.46 (.65)	0.52 (.71)	0.60 (.78)	0.69 (.85)	0.80 (.93)	0.93 (1.00)
0.9	0.24 (.42)	0.28 (.47)	0.32 (.51)	0.36 (.55)	0.40 (.60)	0.46 (.65)	0.53 (.72)	0.63 (.81)	0.80 (.93)

w^*_{OR} contains $\pi_2(1-\pi_2)$ in the numerator. The range of RCE values for $\gamma = 5$ is between 0.59 and 0.96, and the range for $\gamma = 2$ is between 0.77 and 1.00. For estimating OR, the lowest value of $RCE_x(0.5)$, 0.59, occurs when $\gamma = 5$, $\pi_1 = 0.5$, and $\pi_2 = 0.1$ or 0.9.

Tables 1–3 show that the efficiencies of a balanced design can vary substantially depending on whether the treatment effect measure is RD, RR, or OR, the value of the cost ratio γ, and obviously also on the values of $\theta^T = (\pi_1, \pi_2, \rho_1, \rho_2)$. Because θ and the cost ratio γ can vary in many different ways, it can be difficult to discern general trends and patterns as one or more of these parameters vary unless we vary only one of the parameters and fix the rest. For example, consider the effect on the optimal allocation w^*_x when all parameters are fixed except the value of only one parameter in the following order: γ, ρ_1, ρ_2, π_1, and π_2. From the tables and formula for w^*_x, we observe that if all other parameters are fixed, then w^*_x is (i) a decreasing function of γ, (ii) an increasing function of ρ_1, (iii) a decreasing function of ρ_2. Further, for the treatment effect measure RR, w^*_x is a decreasing function of π_1, for RD, it is an increasing function of π_1 until 0.5 after which it decreases, and for OR, it is a decreasing function of π_1 until 0.5 after which it increases. As a function of π_2, we observe an opposite trend for RD, RR, and OR. The R code available

Table 3

RCE of a balanced design ($w = 0.5$) compared to the optimal design for estimating OR with fixed number of clusters under different combinations of π_1 and π_2 when $\rho_1 = 0.05$, $\rho_2 = 0.1$, $m = 20$, and $\gamma = 5(2)$

π_1/π_2	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
0.1	0.80 (.93)	0.90 (.98)	0.94 (1.00)	0.95 (1.00)	0.96 (1.00)	0.95 (1.00)	0.94 (1.00)	0.90 (.98)	0.80 (.93)
0.2	0.68 (.85)	0.80 (.93)	0.85 (.96)	0.87 (.97)	0.88 (.98)	0.87 (.97)	0.85 (.96)	0.80 (.93)	0.68 (.85)
0.3	0.63 (.80)	0.75 (.89)	0.80 (.93)	0.83 (.95)	0.83 (.95)	0.83 (.95)	0.80 (.93)	0.75 (.89)	0.63 (.80)
0.4	0.60 (.78)	0.72 (.87)	0.77 (.91)	0.80 (.93)	0.81 (.94)	0.80 (.93)	0.77 (.91)	0.72 (.87)	0.60 (.78)
0.5	0.59 (.77)	0.71 (.87)	0.77 (.91)	0.79 (.92)	0.80 (.93)	0.79 (.92)	0.77 (.91)	0.71 (.87)	0.59 (.77)
0.6	0.60 (.78)	0.72 (.87)	0.77 (.91)	0.80 (.93)	0.81 (.94)	0.80 (.93)	0.77 (.91)	0.72 (.87)	0.60 (.78)
0.7	0.63 (.80)	0.75 (.89)	0.80 (.93)	0.83 (.95)	0.83 (.95)	0.83 (.95)	0.80 (.93)	0.75 (.89)	0.63 (.80)
0.8	0.68 (.85)	0.80 (.93)	0.85 (.96)	0.87 (.97)	0.88 (.98)	0.87 (.97)	0.85 (.96)	0.80 (.93)	0.68 (.85)
0.9	0.80 (.93)	0.90 (.98)	0.94 (1.00)	0.95 (1.00)	0.96 (1.00)	0.95 (1.00)	0.94 (1.00)	0.90 (.98)	0.80 (.93)

in Web Appendix A.4 allows the user to generate the RCEs and w_x^* based on the optimal allocation for any different sets of values for the parameters.

3. Maximin Optimal Design

In Section 2, we derived the optimal allocation w_x^* for a particular measure $x \in \{RD, RR, OR\}$ when θ is assumed known. Clearly, the optimal allocation w_x^* depends on the vector of parameters $\theta^T = (\pi_1, \pi_2, \rho_1, \rho_2)$, the cluster size m and the cost ratio γ, and so they are termed locally optimal designs (Chernoff, 1953). In practice, the cluster size and cost ratio are likely known before the study, but the values of π_1, π_2, ρ_1, and ρ_2 are not. Consequently, nominal values for those parameters are needed before the optimal design can be determined. But if the parameters are mis-specified and take different values in the actual trial, then the selected design can end up being far from optimal.

A maximin optimal design can guard against this risk. In general, a maximin optimal design is a design that maximizes some measure of performance in the worst case scenario when larger values of the measure are more desirable, see for example, Dette and Biedermann (2003) or Biedermann, Dette, and Pepelyshev (2004). In our context, we chose to maximize the RCE in the worst case scenario. Conceptually, the maximin optimal design can be found as follows: (i) Specify plausible ranges of values for unknown parameters; (ii) For each design (for each fixed w in our case), find the worst configuration within the set of possible parameter values, that is, the one that gives the smallest RCE; then (iii) Select the design (value of w) that maximizes the smallest RCE. This design is the maximin optimal design.

To find the maximin optimal design for a two-arm CRT with binary outcomes with cost consideration, we proceed as follows. First, specify a plausible region Θ containing all plausible values of θ. We seek the allocation scheme that maximizes the minimum RCE that can arise so long as θ is in the user-specified region Θ. More formally, our design criterion is to find maximin optimal proportion of clusters to assign to arm 1, $w_x^{m*} \in (0, 1)$, such that $min(RCE_x(\theta, w, m, \gamma)|\theta \in \Theta)$ is maximized. To this end, recall that $d_h = 1 + (m-1)\rho_h$, $h \in \{1, 2\}$ and let

$$y_x(\theta) = \begin{cases} \dfrac{\pi_2(1-\pi_2)}{\pi_1(1-\pi_1)}\dfrac{d_2}{d_1} & \text{if } x = RD \\[2ex] \dfrac{\pi_1(1-\pi_2)}{\pi_2(1-\pi_1)}\dfrac{d_2}{d_1} & \text{if } x = RR \\[2ex] \dfrac{\pi_1(1-\pi_1)}{\pi_2(1-\pi_2)}\dfrac{d_2}{d_1} & \text{if } x = OR. \end{cases}$$

The quantity $y_x(\theta)$ does not have a meaningful interpretation but it allows us to write the above expressions for the three measures w_{RD}^*, w_{RR}^*, and w_{OR}^* more succinctly as

$$w_x^* = \frac{1}{1 + \sqrt{\gamma y_x(\theta)}},$$

where, as before, $x \in \{RD, RR, OR\}$. It also provides a means of translating the four ranges for the four parameters into a single overall range. For the given Θ, let $\underline{y_x} = min(y_x(\theta))$ and let $\overline{y_x} = max(y_x(\theta))$, where the optimization is over the plausible region Θ. These are important quantities needed to obtain the maximin allocation rule. For example, if the treatment measure is OR, $m = 20$, $0.3 \le \pi_1 \le 0.5$, $0.2 \le \pi_2 \le 0.3$, $0.1 \le \rho_1 \le 0.2$ and $0.1 \le \rho_2 \le 0.2$, we have $\underline{y_{OR}} = 0.604$ and $\overline{y_{OR}} = 2.586$. We show in Web Appendix A.1 that the maximin optimal proportion of clusters to assign to arm 1 in a two-arm CRT is

$$w_x^{m*} = \frac{(\sqrt{\gamma} + \sqrt{\underline{y_x}})^2 - (\sqrt{\gamma} + \sqrt{\overline{y_x}})^2}{(\sqrt{\gamma} + \sqrt{\overline{y_x}})^2(\underline{y_x} - 1) - (\sqrt{\gamma} + \sqrt{\underline{y_x}})^2(\overline{y_x} - 1)}. \quad (1)$$

For the same illustrative example, a direct calculation shows $w_{OR}^{m*} = 0.386$ if $\gamma = 2$ and $w_{OR}^{m*} = 0.473$ if $\gamma = 1$. The practical implication is that if cost in arm 1 is twice as expensive as that for arm 2, the optimal maximin strategy for the given plausible region is to allocate about 10% fewer subjects to the more expensive arm.

It is interesting to note that the optimal allocation rule has the same form for all three measures, RD, RR, and OR, but the optimal proportion of clusters to assign to arm 1 varies because the value of w_x^{m*} depends on $\underline{y_x}$ and $\overline{y_x}$ which depend on the measure of interest. When $\rho_1 = \rho_2$, the formula for w_x^{m*} simplifies and becomes the optimal allocation to arm 1 in an IRT.

Now that we have moved from specifying single values of parameters to specifying ranges of parameters, it is natural to ask how the optimal design depends on the specified range. Table 4 provides examples of how different ranges of ρ_1 and ρ_2 affect the maximin optimal allocation w_x^{m*} for the three measures when π_1 and π_2 are fixed. For all measures, the value of w_x^{m*} increases as ρ_1 increases and as ρ_2 decreases. This is the similar to the result in Section 2 in which the value of w_x^* increases as ρ_1 increases and ρ_2 decreases. The maximal optimal design allows specifying the ranges of ρ_1 and ρ_2 instead of single values of ρ_1 and ρ_2, but the maximal optimal allocation w_x^{m*} depends on the locations of those ranges. Limited by space, we do not show examples of how different ranges of π_1 and π_2 affect the maximin optimal allocation w_x^{m*}. Web Appendix A.4 contains R code for calculating w_x^{m*} for a user-specified parameter set Θ. Interested readers can use the code to further explore the effects of ranges of parameters on the maximin optimal design.

The locally optimal design in Section 2 is for a particular point in the set Θ. The maximin optimal design is unique and a globally optimal design, which considers the worst case scenario that can arise within the set of plausible values of $\theta \in \Theta$. It can be shown that w_x^{m*} is a locally optimal design for a point in the set Θ, and the RCE of the maximin optimal design is 1 when that particular point is the true value of θ. This is a common feature of maximin optimal designs in general, see for example, the discussion in Dette and Biedermann (2003).

4. Guidance for Constructing a Maximin Optimal Design for CRTs and Example

We now provide a step by step approach to find a maximin optimal design for a two-arm CRT with a binary outcome

Table 4

Maximin optimal allocation w_x^{m*} for outcome measure x and different ranges of ρ_1 and ρ_2 when $\pi_1 \in [0.3, 0.5]$, $\pi_2 \in [0.2, 0.3]$, $\gamma = 5$, and $m = 20$

	ρ_1/ρ_2	[0, 0.1]	[0.1, 0.2]	[0.2, 0.3]
$x = RD$				
	[0, 0.1]	0.315	0.247	0.212
	[0.1, 0.2]	0.408	0.327	0.285
	[0.2, 0.3]	0.461	0.375	0.330
$x = RR$				
	[0, 0.1]	0.226	0.175	0.150
	[0.1, 0.2]	0.297	0.233	0.201
	[0.2, 0.3]	0.341	0.271	0.235
$x = OR$				
	[0, 0.1]	0.273	0.210	0.179
	[0.1, 0.2]	0.358	0.281	0.243
	[0.2, 0.3]	0.408	0.326	0.283

when the total number of clusters is fixed in advance.

Step 1. Estimate the cluster size m and the cost ratio γ of the total cost per cluster in arm 1 compared to arm 2. In our maximin optimal design method, these are assumed known. Like many design methods for CRTs, our method assumes cluster size is constant. If there is some uncertainty about the value of m or γ, a sensitivity analysis can be conducted, varying these values.

Step 2. Select a treatment effect measure. As mentioned previously, the maximin optimal design can be different for the different measures of treatment effect for binary outcomes, the RD, the RR and the OR. For the design, investigators should use the treatment effect measure that they plan to estimate, as specified in the study protocol. For example, if the protocol calls for using a mixed logistic regression model, investigators should select the OR as their measure for the design work.

Step 3. Specify ranges of possible values for the parameters $(\pi_1, \pi_2, \rho_1, \rho_2)$. Investigators need to specify minimum and maximum values for plausible success rates and ICCs in each condition. Previous studies, pilot data, and expert opinion can help to specify these ranges. There is a large literature on elicitation of prior distributions for parameters in Bayesian analyzes; see, for example, Garthwaite, Kadane, and O'Hagan (2005). The task here is easier than soliciting a prior distribution, since we need only a range for each parameter, not a full joint probability distribution. However, some ideas for specifying parameter locations and intervals can be applied. One may ask the question "What is the range of values within which the response rate will have a 95% chance to occur?" to solicit a 95% credible interval for a parameter. The range for each of the two ICCs may be harder to elicit, since the ICC is a less intuitive parameter than the success rate, but there are an increasing number of literature reviews summarizing ICC values for various types of studies (for e.g., Hade et al., 2010; Crespi, Maxwell, and Wu, 2011), and these can help provide information for specifying a plausible range for each of the ICCs.

Step 4. Compute the maximin optimal allocation w_x^{m*} and assign this proportion of clusters to arm 1 and the remainder to arm 2. More precisely, for a fixed total number of clusters k, the optimal number of clusters to assign to arm 1 is kw_x^{m*}, rounded to the nearest integer.

We now apply the maximin approach to redesign a CRT for the Samoan Women's Health Study (Mishra et al., 2007) to illustrate these steps. This study used a cluster randomized design for an intervention trial whose objective was to increase mammography usage among Samoan American women. A total of 61 Samoan-language churches in two counties in southern California agreed to participate in the study, providing our fixed total k. Churches served as clusters and were randomly assigned to either participate in a culturally tailored breast cancer education program or a control condition. The intervention included specially developed English and Samoan language breast cancer educational booklets, skill building and behavioral exercises, and interactive group discussion sessions. In the control arm, women were provided with standard breast cancer educational materials. The mean cluster size was 14 and we use this value as the constant cluster size. The binary outcome was self-reported mammogram use at follow-up. Because the intervention condition required substantially more resources than the control condition, our estimation was that a cost ratio of $\gamma = 5$ was justified.

Next, we consider specifying the range of possible values for each of the parameters π_1, π_2, ρ_1, and ρ_2. An earlier study reported prevalences of mammography use of 0.224 and 0.244 among Samoan women in Hawaii and Los Angeles, respectively (Mishra, Luce, and Hubbell, 2001). Treating this as an estimate for the proportion of mammography use by Samoan women in the control condition, we specify the range of values for π_2 as [0.2, 0.3]. To estimate a possible range of values for the proportion of responders in the intervention arm, one may proceed as follows. First, we believe the intervention will increase mammography use and so the smallest value of π_1 should be larger than the largest anticipated value of π_2. Second, we have less certainty about the intervention effect, so we specify a wider range of possible values for π_1. Accordingly, we set the range of π_1 to be [0.3, 0.6]. The next task is to specify reasonable ranges for the ICCs. This is always problematic when no similar prior studies are available, which is the case here. We combed the literature and found that Hade et al. (2010) had reported ICCs for cancer screening CRTs ranged from 0.05 to 0.3. However, not all of the clusters were churches and not all of the trials involved mammography use. Nevertheless, given the limited information available, we worked with these ranges of values for both ρ_1 and ρ_2. For illustration purposes, we also consider the case when the cost ratio is $\gamma = 2$ to ascertain whether the maximin optimal design is sensitive to the cost ratio value.

Table 5
Maximin optimal designs for the Samoan Women's Health Study with 61 clusters and 14 subjects per cluster

		RD	RR	OR
$\gamma = 2$				
	w_x^{m*}	0.430	0.316	0.382
	k_1	26	19	23
	k_2	35	42	38
$\gamma = 5$				
	w_x^{m*}	0.315	0.210	0.272
	k_1	19	13	17
	k_2	42	48	44

Results, obtained using formula (1), are summarized in Table 5. The numbers of clusters have been rounded to the nearest integer. Recalling that the cost ratio γ is the total cost in arm 1 relative to arm 2, we observe that in general, fewer clusters are allocated to the more costly arm 1. We also see that the maximin optimal design is indeed sensitive to the cost ratio value; for example, for RD, the number of clusters allocated to arm 1 decreases from 26 to 19 as γ is changed from 2 to 5.

Let us compare the RCE of our maximin optimal design to the RCE of the balanced design for each measure. We first consider RD. Figure 1a shows RCEs of the maximin optimal design and the balanced design when the cost ratio is 2, which is relatively small. The quantity on the x-axis is $y_{RD}(\theta)$, which, we recall, does not have a meaningful interpretation but does serve to translate the four parameter ranges into one overall range. For the Samoan Women's Health study, for RD, the minimum value of $y_{RD}(\theta)$, which is 0.22, occurs when $(\pi_1, \pi_2, \rho_1, \rho_2) = (0.5, 0.2, 0.3, 0.05)$ and the maximum value, which is 2.97, occurs when $(\pi_1, \pi_2, \rho_1, \rho_2) = (0.3, 0.3, 0.05, 0.3)$. We observe that over the whole range of $y_{RD}(\theta)$, the lowest RCE for the maximin optimal design is about 0.91, while the lowest RCE for the balanced design is about 0.83. In addition, for a larger portion of the range of $y_{RD}(\theta)$, the RCE of the maximin optimal design is larger than that of the balanced design.

Figure 1b shows results for cost ratio $\gamma = 5$. We observe that the RCE of the maximin optimal design is always larger than 0.92, while the RCEs of balanced designs can be as low as 0.66. In addition, the RCE of the maximin optimal design exceeds than that of the balanced design for almost the whole range of $y_{RD}(\theta)$, suggesting that the maximin optimal design greatly outperforms the balanced design when the cost ratio is 5.

Figure 1c and 1d show RCEs for the outcome measure RR when the cost ratio is $\gamma = 2$ and 5. Here, $y_{RR}(\theta)$ ranges from about 0.2–18. From both plots, we observe that the maximin optimal design outperforms the balanced design over almost the entire range of possible parameter values. The lowest RCEs of the balanced designs are less than 0.6 and 0.4 for $\gamma = 2$ and $\gamma = 5$, respectively, and both are lower than those in Figure 1a and 1b . This suggests that if the outcome measure is RR rather than RD, the performance of the balanced design is more sensitive to mis-specified parameters and the maximin optimal design is more helpful to avoid low RCE whether the cost in the intervention arm is twice or five times that in the control arm.

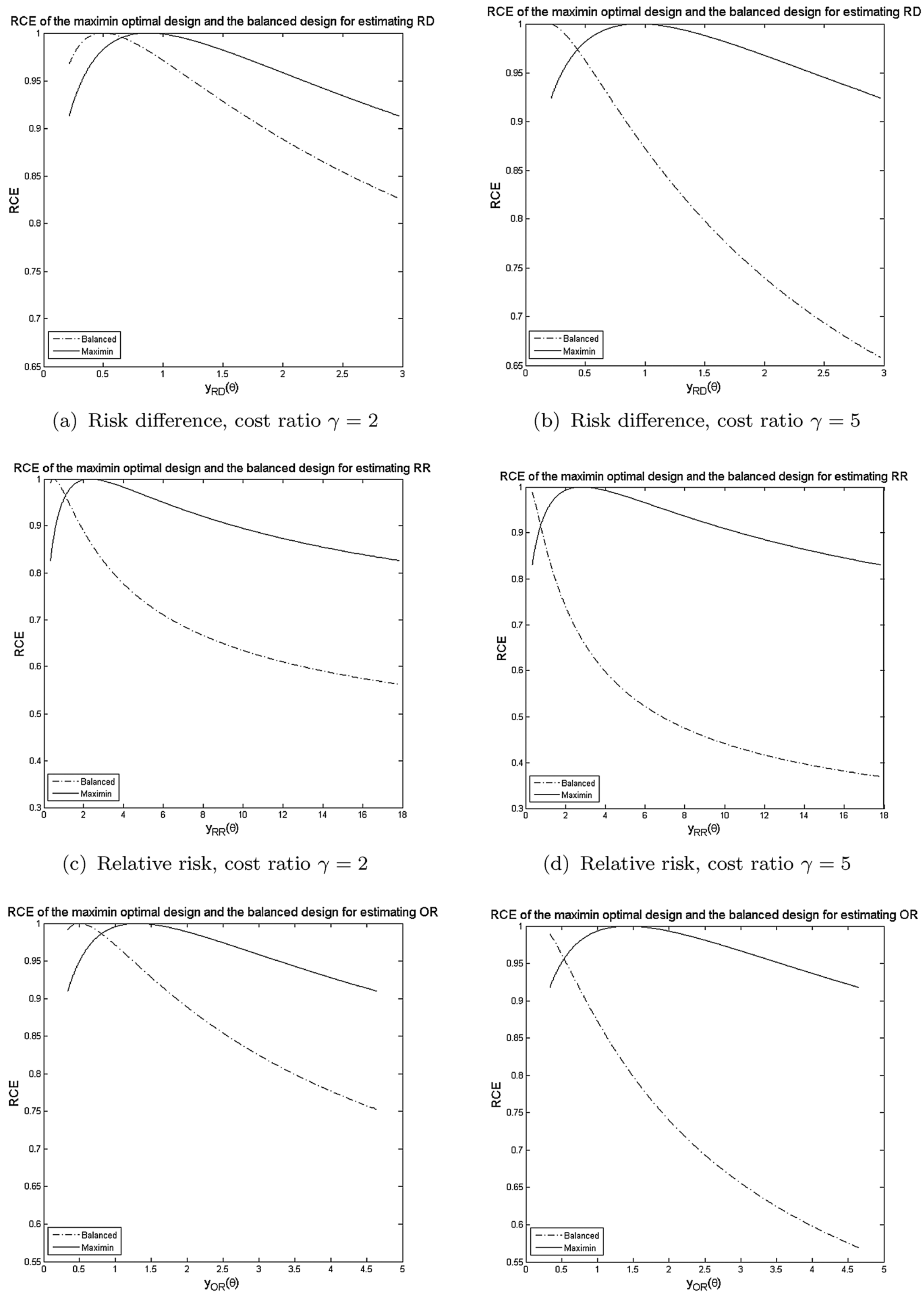
Figure 1e and 1f shows RCEs for OR for the cost ratios $\gamma = 2$ and 5. The lowest RCEs of the maximin design are larger than 0.90, and it clearly outperforms the balanced design for almost all possible parameter values. Note that the lowest RCEs of the balanced design are about 0.75 and 0.57 for $\gamma = 2$ and $\gamma = 5$, respectively, and both are lower than those for RD but larger than those for RR. The implication is that the balanced design is less sensitive for estimating OR than for estimating RR but more sensitive than estimating RD. The upshot is that the maximin optimal design is again helpful to avoid having a low RCE.

In the Samoan Women's Health Study, the planned outcome analysis involved estimating the OR using generalized estimating equations (GEE). According to Table 5, the maximin allocation value is 0.272 and the maximin optimal design would allocate 17 churches to the intervention condition and 44 to the control condition. From Figure 1f, we see that the maximin design does an excellent job of guarding against low RCE. The maximin optimal design is generally more efficient (i.e., has a larger RCE) than the balanced design and is robust to mis-specifications of the ICCs and the success rates in the two arms. While some investigators may not be comfortable with such an unequal allocation and may prefer to adjust it, this information can be useful as part of the overall study planning process and can lead to designs that are superior to a default balanced design.

5. Discussion

Much of the research in finding optimal allocation schemes for a CRT involve locally optimal designs in which the design depends on the success rates and ICCs, which are typically unknown in advance. Such single best guesses for these parameters can result in substantial loss in efficiency if these parameters are mis-specified. In this article, we provide a novel approach to designing a two-arm CRT that allows a range of plausible values to be specified for each of the design parameters. The approach is flexible and applies when the intervention effect is measured in terms of RD, RR, or OR. We provide closed form formulae for the optimal proportions of equal-sized clusters in the two arms for three common outcome measures when we have a predetermined fixed number of clusters. Our optimal design maximizes a cost efficiency measure that combines the precision of the estimated intervention effect and cost of the study. We also compare our proposed designs with the popular balanced designs using the RCE measure and show that RCEs of a balanced design can be very low relative to maximin optimal designs.

We consider three treatment effect measures, RD, RR, and OR, in our work. RR is often used in randomized controlled trials and cohort studies and OR is typically used for cross-sectional and case-control studies (Sistrom, Garvan, and Grobbee, 2011). OR is also used in randomized controlled trials (Knol and Duijnhoven, 2004). Ukoumunne et al (2008) discussed how these measures can affect the results using the GEE method of analysis. Because the same design can have



(a) Risk difference, cost ratio $\gamma = 2$

(b) Risk difference, cost ratio $\gamma = 5$

(c) Relative risk, cost ratio $\gamma = 2$

(d) Relative risk, cost ratio $\gamma = 5$

(e) Odds ratio, cost ratio $\gamma = 2$

(f) Odds ratio, cost ratio $\gamma = 5$

Figure 1. RCE of the maximin optimal design (solid line) and the balanced design (dashed line) for estimating the RD, RR, and OR for cost ratios $\gamma = 2$ and $\gamma = 5$ when $\pi_1 \in [0.3, 0.6]$, $\pi_2 \in [0.2, 0.3]$, $\rho_1 \in [0.05, 0.3]$, $\rho_2 \in [0.05, 0.3]$, and all clusters have $m = 14$ subjects.

different efficiencies under different outcome measures, investigators should ensure that they use the same measure for their design and their analysis.

Throughout, we have assumed that the cluster size is constant. In practice, cluster sizes often vary. Several complications arise when adapting design methods for CRTs with equal cluster size to CRTs with unequal cluster sizes. For example, in the latter case, there is now a choice of several different weighting schemes for computing treatment effect estimates and several different variance estimators just for RD alone (Kerry and Bland, 2001; Guittet, Ravaud, and Giraudeau, 2006). Since the optimal design depends on the specific estimators, deriving, and studying formulas for maximin optimal designs for CRTs with varying cluster sizes for the various risk measures and weighting schemes would require a substantial effort. Compounding the issue is that there is currently no agreed upon method for designing a CRT with unequal cluster sizes in the literature and so it is not clear how to fairly evaluate performance of our proposed maximin optimal designs when cluster sizes are unequal. However, we can offer some general observations for designing a maximin optimal CRT with unequal cluster sizes.

First, we explored how the maximin optimal allocation varies as a function of the common cluster size for a range of scenarios. The results in Web Appendix A.2 show that the maximin allocation strategy generally varies very little as the common cluster size is changed, except when cluster sizes are small for some scenarios. This provides some assurance that the method may work acceptably in many settings.

We were also able to find a result from Kang et al. (2005) that seems helpful. They worked on sample size issues for detecting a user-specified RD and derives the design effect for $Var(\hat{\pi})$ under varying cluster sizes when equal weights for subjects (that is, weights equal to cluster size) are assumed. The modified design effect has the formula $1 + [E(M) - 1]\rho + E(M)\rho CV^2$, where $E(M)$ is the expected cluster size and CV is the coefficient of variation of cluster size. We amended our optimal allocation formula for RD to use this design effect. Plots in the Web Appendix A.3 show how the ratio of the optimal allocation for varying cluster size to the optimal allocation for constant cluster size varies as a function of the CV for the RD measure for selected scenarios. A ratio of 1 indicates that the optimal allocations are the same for both formulas. When CV=0, we have constant cluster size and the two formulas coincide. As the CV is increased to 0.8, we observe an increase or decrease of only about 4% in the ratio, which is unlikely to make much difference after we round an allocation to whole numbers of clusters. In practice, the CV for cluster sizes is rarely larger than 0.8 (Eldridge et al., 2006; Carter, 2010). These figures also show that varying the expected cluster size is likely to have little impact.

While we are unable to fully explore the impact of varying cluster sizes on the maximin optimal allocation analytically, the observations suggest that using the mean cluster size in our formulas for CRTs with constant cluster size may produce acceptable results in some settings. Researchers may wish to conduct similar sensitivity analyzes for their particular user-specified settings.

We conclude by noting that there are alternative design approaches when there are unknown parameters in the model. One option is to use a Bayesian approach where we first solicit a joint prior distribution for all parameters and then find the optimal proportion to arm 1 (or arm 2) that minimizes the averaged expected variance with respect to the joint prior distribution of π_1, π_2, ρ_1, and ρ_2. Frequently, the priors for the various parameters are assumed to be independent. Interestingly, while there is work on analyzing binary outcome in IRTs using Bayesian methods (Matthews, 1999), we were unable to find papers that focus on constructing Bayesian optimal designs for CRTs. One reason may be the practical difficulties encountered in eliciting a joint prior distribution for ICCs and the response rates.

In summary, the maximin method proposed in this work may appear technically more complex but may actually be simpler to implement in practice because it is relatively easy to elicit a range of plausible values for each of the parameters in the design problem. Additionally, the maximin optimal design offers some protection against the worst case scenario and is generally more robust than locally optimal designs. Our results also show they tend to be more efficient than balanced designs in terms of the RCE measure. Other design work in a non-CRT setting also supports such a conclusion when a maximin (or equivalently a minimax) optimal design was used to estimate parameters in a nonlinear regression model, see for example, King and Wong (2000), Tan, and Berger (2002), Dette and Biedermann (2003), Biedermann et al. (2004), Tekle, Tan, and Berger (2008), and Ouwens, Rodriguez, Ortiz, and Martnez (2014).

6. Supplementary Materials

Web Appendices referenced in Sections 2, 3, and 5, which include R code for implementing the methods, are available with this article at the *Biometrics* website on Wiley Online Library.

Acknowledgments

The research of Professor Wong and Wu reported in this article was partially supported by the National Institute of General Medical Sciences of the National Institutes of Health under the Grant Award Number R01GM107639. The research of Professor Crespi was partially supported by grants CA16042 and TR000124 from the National Institute of Health. Professor Wong wishes to thank the hospitality at the Isaac Newton Institute of Mathematical Sciences at Cambridge where he worked on this manuscript during the 1-week health care design workshop hosted by Professor Rosemary Bailey in July 2015. Wu is also grateful to the International Biometric Society, Western North American Region 2015 Student Paper Award Committee for receiving a distinguished paper award in the student paper competition based on an earlier version of this article. The contents in this article are solely the responsibility of the authors and do not necessarily represent the official views of the National Institutes of Health. The authors would like to thank all members of the editorial team for providing very helpful and valuable feedback on an earlier version of the article.

References

Biedermann, S., Dette, H., and Pepelyshev, A. (2004). Maximin optimal designs for a compartmental model. *mODa 7 - Advances in Model-Oriented Design and Analysis*, 41–49. Springer-Verlag Berlin Heidelberg New York.

Breukelen, G. J. and Candel, M. J. (2012). Calculating sample sizes for cluster randomized trials: we can keep it simple and efficient. *Journal of Clinical Epidemiology* **65**, 1212–1218.

Brittain, E. and Schlesselman, J. J. (1982). Optimal allocation for the comparison of proportions. *Biometrics* **38**, 1003–1009.

Carter, B. (2010). Cluster size variability and imbalance in cluster randomized controlled trials. *Statistics in Medicine* **29**, 2984–2993.

Chernoff, H. (1953). Locally optimal designs for estimating parameters. *The Annals of Mathematical Statistics* **24**, 586–602.

Crespi, C. M., Maxwell, A. E., and Wu, S. (2011). Cluster randomized trials of cancer screening interventions: are appropriate statistical methods being used? *Contemporary Clinical Trials* **32**, 477–484.

Crespi, C. M., Wong, W. K., and Mishra, S. (2009). Using second-order generalized estimating equations to model heterogeneous intraclass correlation in cluster randomized trials. *Statistics in Medicine* **28**, 814–827.

Crespi, C. M., Wong, W. K., and Wu, S. (2011). A new dependence parameter approach to improve the design of cluster randomized trials with binary observations. *Clinical Trials* **8**, 687–698.

Dette, H. (2004). On robust and efficient designs for risk estimation in epidemiological studies. *Scandinavian Journal of Statistics.* **31**, 319–331.

Dette, H. and Biedermann, S. (2003). Robust and efficient designs for the Michaelis–Menten model. *Journal of the American Statistical Association* **98**, 679–686.

Donner, A. and Klar, N. (2000). *Design and Analysis of Cluster Randomization Trials in Health Research*. New York, NY: Oxford University Press.

Donner, T., Birkett, N., and Buck, C. (1981). Randomization by cluster: Sample size requirements and analysis. *American Journal of Epidemiology.* **114**, 906–914.

Eldridge, S. M., Ashby, D., and Kerry, S. (2006). Sample size for cluster randomized trials: The effect of coefficient of variation of cluster size and analysis method. *International Journal of Epdeimiology* **35**, 1292–1300.

Eldridge, S. M., Ukoumunne, O. C., and Carlin, J. B. (2009). The intra-cluster correlation coefficient in cluster randomized trials: A review of definitions. *International Statistical Review* **77**, 378–394.

Gail, M. H., Mark, S. D., Carroll, R. J., Green, S. B., and Pee, D. (1996). On design considerations and randomization-based inference for community intervention trials. *Statistics in Medicine* **15**, 1069–1092.

Garthwaite, P. H., Kadane, J. B., and O'Hagan, A. (2005). Statistical methods for eliciting probability distributions. *Journal of the American Statistical Association* **100**, 680–700.

Guittet, L., Ravaud, P., and Giraudeau, B. (2006). Planning a cluster randomized trial with unequal cluster sizes: Practical issues involving continuous outcomes. *BMC Medical Research Methodology* **6**, 17.

Hade, E. M., Murray, D. M., Pennell, M. L., Rhoda, D., Paskett, E. D., Champion, V. L., et al. (2010). Intraclass correlation estimates for cancer screening outcomes: Estimates and applications in the design of group-randomized cancer screening studies. *Journal of the National Cancer Institute* **40**, 97–103.

Hayes, R. J. and Moulton, L. H. (2009). *Cluster Randomised Trials.* Boca Ration, FL: CRC Press.

Kang, S.-H., Ahn, C. W., and Jung, S.-H. (2003). Sample size calculation for dichotomous outcomes in cluster randomization trials with varying cluster size. *Drug Information Journal* **37**, 109–114.

Kerry, S. M. and Bland, J. M. (2001). Unequal cluster sizes for trials in English and Welsh general practice: Implications for sample size calculations. *Statistics in Medicine* **20**, 377–390.

King, J. and Wong, W. K. (2000). Minimax *D*-optimal designs for the logistic model. *Biometrics* **56**, 1263–1267.

Knol, N. J. and Duijnhoven, R. J. (2004). Proportions, odds and risk. *Radiology* **230**, 12–19.

Lubin, J. H. (1980). Some efficency comments on group size in study design. *American Journal of Epidemiology* **111**, 347–359.

Matthews, J. N. S. (1999). Effect of prior specification on Bayesian design for two-sample comparison of a binary outcome. *The American Statistician* **53**, 254–256.

Meydrich, E. F. (1978). Cost considerations and sample size requirements in cohort and case-control studies. *Scandinavian Journal of Statistics* **107**, 201–205.

Milliken, G. A. and Johnson, D. E. (1984). *Analysis of Messy Data Volume I: Designed Experiments*. Belmont, CA: Lifetime Learning Publications.

Mishra, S. I., Bastani, R., Crespi, C. M., Chang, L. C., Luce, P. H., and Baquet, C. R. (2007). Results of a randomized trial to increase mammogram usage among Samoan women. *Cancer Epidemiology, Biomarkers and Prevention* **16**, 2594–2604.

Mishra, S. I., Luce, P. H., and Hubbell, F. A. (2001). Breast cancer screening among American Samoan women. *Preventive Medicine* **33**, 9–17.

Moerbeek, M., Breukelen, G. J., and Berger, M. P. (2000). Design issues for experiments in multilevel populations. *Journal of Educational and Behavioral Statistics* **25**, 271–284.

Moerbeek, M. and Teerenstra, S. (2016). *Power Analysis of Trials with Multilevel Data*. Boca Raton, FL: CRC Press.

Morgenstern, H. and Winn, D. M. (1983). A method for determining the sampling ratio in epidemiologic studies. *Statistics in Medicine* **2**, 387–396.

Ouwens, M. J. N. M., Tan, F. E. S., and Berger, M. P. F. (2002). Maximin D-optimal designs for longitudinal mixed effects models. *Biometrics* **58**, 735–741.

Raudenbush, S. W. (1997). Statistical analysis and optimal design for cluster randomized trials. *Psychological Methods* **2**, 173–185.

Raudenbush, S. W. and Liu, X. (2000). Statistical power and optimal design for multisite randomized trials. *Psychological Methods* **5**, 199–231.

Ridout, M. S., Demetrio, C. G. B., and Firth, D. (1999). Estimating intraclass correlation with binary data. *Biometrics* **55**, 137–148.

Rodriguez, C., Ortiz, I., and Martnez, I. (2014). Locally and maximin optimal designs for multi-factor nonlinear models. *Statistics: A Journal of Theoretical and Applied Statistics* **49**, 1157–1168.

Sistrom, C. L., Garvan, C. W., and Grobbee, D. E. (2011). Potential misinterpretation of treatment effects due to use of odds ratios and logistic regression in randomized controlled tirals. *PLoS ONE* **6**, e21248.

Tekle, F. B., Tan, F. E. S., and Berger, M. P. F. (2008). Maximin D-optimal designs for binary longitudinal responses. *Computational Statistics & Data Analysis* **52**, 5253–5262.

Ukoumunne, O. C., Forbes, B., Carlin, J. B., and Gulliford, M. C. (2008). Comparison of the risk difference, risk ratio and odds ratio scales for quantifying the unadjusted intervention effect in cluster randomized trials. *Statistics in Medicine.* **27**, 5143–5155.

Wu, S., Crespi, C. M., and Wong, W. K. (2012). Comparison of methods for estimating the intraclass correlation coefficient for binary responses in cancer prevention cluster randomized trials. *Contemporary Clinical Trials* **33**, 869–880.

Yanagawa, T. and Bolt, W. J. (1977). Optimal sampling ratios for prospective studies. *American Journal of Epidemiology* **106**, 436–437.

Received August 2015. *Revised April 2016 and December* 2016. *Accepted December* 2016.