Received: 21 June 2018 | Revised: 6 March 2019 | Accepted: 9 March 2019
DOI: 10.1002/sim.8153

RESEARCH ARTICLE

WILEY Statistics in Medicine

Optimal designs in three-level cluster randomized trials with a binary outcome

Jingxia Liu[1,2] | Lei Liu[2] | Graham A. Colditz[1]

[1]Division of Public Health Sciences, Department of Surgery, Washington University School of Medicine, St. Louis, Missouri

[2]Division of Biostatistics, Washington University School of Medicine, St. Louis, Missouri

Correspondence
Jingxia Liu, Division of Public Health Sciences, Department of Surgery, Washington University School of Medicine, St. Louis, MO 63110; or Division of Biostatistics, Washington University School of Medicine, St. Louis, MO 63110.
Email: esther@wustl.edu

Funding information
Siteman Cancer Center at Washington University School of Medicine and Barnes-Jewish Hospital, Grant/Award Number: P30 CA91842; Washington University Institute of Clinical and Translational Sciences, Grant/Award Number: UL1TR000448; National Center for Advancing Translational Sciences (NCATS) of the National Institutes of Health (NIH)

Cluster randomized trials (CRTs) were originally proposed for use when randomization at the subject level is practically infeasible or may lead to a severe estimation bias of the treatment effect. However, recruiting an additional cluster costs more than enrolling an additional subject in an individually randomized trial. Under budget constraints, researchers have proposed the optimal sample sizes in two-level CRTs. CRTs may have a three-level structure, in which two levels of clustering should be considered. In this paper, we propose optimal designs in three-level CRTs with a binary outcome, assuming a nested exchangeable correlation structure in generalized estimating equation models. We provide the variance of estimators of three commonly used measures: risk difference, risk ratio, and odds ratio. For a given sampling budget, we discuss how many clusters and how many subjects per cluster are necessary to minimize the variance of each measure estimator. For known association parameters, the locally optimal design is proposed. When association parameters are unknown but within predetermined ranges, the MaxiMin design is proposed to maximize the minimum of relative efficiency over the possible ranges, that is, to minimize the risk of the worst scenario.

KEYWORDS
cluster randomized trial (CRT), dissemination and implementation, generalized estimating equation (GEE), intracluster correlation coefficient (ICC), nested correlation structure

1 | INTRODUCTION

Cluster randomized trials (CRTs) were originally proposed for use when randomization at the subject level is practically infeasible or may possibly lead to a severe estimation bias of the treatment effect. In practice, implementation strategies play important roles in dissemination and implementation research. Compared to subject-level randomized trials, CRT designs have appealing features for implementation science in public health and clinical medicine. Furthermore, CRTs are greatly needed for effectiveness research from science to practice.[1,2] Therefore, there has been growing interest in the design of CRTs.[3-9] The unit of randomization might be hospitals, clinics, classrooms, etc. Subjects within a cluster are exposed to common factors and tend to share similar characteristics. The degree of such similarity is commonly quantified by the intracluster correlation coefficient (ICC). Recruiting an additional cluster costs more than enrolling an additional subject in an individually randomized trial; thus, researchers have proposed the optimal sample size as a function of sampling costs and the ICC in CRTs.[10-16] "Optimal" means the maximum power and precision for a given sampling budget or the minimum sampling cost for a given power and precision. These approaches show that the optimal sample size depends strongly on the ICC. However, the ICC is usually unknown in CRTs. To overcome this shortcoming, Van Breukelen and Candel considered a range of possible ICC values and presented MaxiMin designs (MMDs) based on relative efficiency (RE) under budget constraints.[17] Wu et al proposed the optimal group allocations for three measures (risk difference [RD], risk ratio [RR], and odds ratio [OR]) in two-level CRTs with binary outcomes through the variances of the maximum likelihood estimators.[18]

CRTs may have three-level structures. For example, subjects in a two-level CRT are measured at different time points. Measurements across the different time points are correlated within a subject, whereas subjects are correlated within a cluster. Another example is that interventions are randomly assigned to medical centers ("practices"), and health care professionals ("providers") within the same practice are trained with the assigned intervention to provide care to participants. Participants could be correlated within a provider, whereas providers could be correlated within a practice. Hereafter, we use a CRT with practice, provider, and participant levels as the three-level example. For simplicity, we consider the same provider size (number of participants from each provider) and equal practice sizes (number of providers per practice).

Generalized estimating equations (GEEs) proposed by Liang and Zeger[19] have been commonly applied to analyze the correlated data in CRTs.[20-24] Liang and Zeger showed that the GEE approach still gives consistent estimates of the regression coefficients provided that the marginal model is correctly specified, even if the working correlation matrix is incorrectly assumed.[19] In this paper, we aim to propose an optimal design (OD) in three-level CRTs, in which "optimal" refers to the minimization of the variance of each measure estimator for a given sampling budget. We assume the nested exchangeable correlation structure[25] throughout and utilize the GEE models in a three-level CRT with a binary outcome. The correlation structure includes correlation among participants within the same provider in the same practice, r, and correlation among participants with different providers in the same practice, ρ. Both r and ρ are assumed to be constant across all practices. Three different link functions in GEE models are considered, eg, identity, log, and logit, where the corresponding regression coefficients are related to RD, RR, and OR, respectively.

For known association parameters (r, ρ), we discuss how many practices m need to be enrolled and how many providers per practice n are sufficient to minimize the variance of each measure estimator under the budget constraints when the provider size K is a predetermined value and K is not a fixed value but within a range $(K_{\min}, K_{\max})$, respectively. This is a locally optimal design (LOD) with corresponding numbers n_{LOD} and m_{LOD}. When the association parameters (r, ρ) are unknown, but we assume that ranges for r and ρ can be obtained from other literature and similar studies, we propose MMDs in the framework of RE to minimize the risk of the worst scenario. RE is defined as the ratio of the variance of each measure estimator for practice size n_{LOD} to n,[17], which is a function of n, r, and ρ. Our goal is to maximize the minimum of RE over the possible ranges of r and ρ.

The organization of this paper is as follows. In Section 2, we briefly summarize the GEE method developed by Liang and Zeger in three-level CRTs,[19] introduce the "nested exchangeable" correlation structure,[25] and derive the variance of the estimator of the treatment for a binary outcome in a two-group comparison. Section 3 presents the LOD for known parameter values under the assumption of a "nested exchangeable" correlation structure. In Section 4, we define the RE and propose MMDs for unknown parameter values of r and ρ. We provide guidance on applying the methods and illustrate using a real CRT, followed by a discussion about the limitations of the proposed approach and directions for future research.

2 | STATISTICAL GEE MODELS IN THREE-LEVEL CRTS

Let Y_{ijk} be a response from participant $k = 1, \dots, K$, for provider $j = 1, \dots, n_i$ in practice $i = 1, \dots, m$. Let $\boldsymbol{X}_{ijk} = (X_{ijk1}, \dots, X_{ijkp})'$ be a covariate vector and $\mu_{ijk} = E(Y_{ijk}|\boldsymbol{X}_{ijk})$ be a marginal mean response given $\boldsymbol{X}_{ijk}$. The marginal model is

$$g(\mu_{ijk}) = \boldsymbol{X}'_{ijk}\boldsymbol{\beta}.$$

Let $\boldsymbol{Y}_{ij} = (Y_{ij1}, \dots, Y_{ijK})$, $\boldsymbol{\mu}_{ij} = (\mu_{ij1}, \dots, \mu_{ijK})$, and $\boldsymbol{X}_{ij} = (\boldsymbol{X}_{ij1}, \dots, \boldsymbol{X}_{ijK})$ be the $1 \times K$ response vector, the $1 \times K$ marginal mean response vector, and the $p \times K$ covariate matrix of provider j in practice i, respectively. Let $\boldsymbol{Y}_i = (\boldsymbol{Y}_{i1}, \dots, \boldsymbol{Y}_{\text{in}_i})'$, $\boldsymbol{\mu}_i = (\boldsymbol{\mu}_{i1}, \dots, \boldsymbol{\mu}_{\text{in}_i})'$, and $\boldsymbol{X}_i = (\boldsymbol{X}_{i1}, \dots, \boldsymbol{X}_{in_i})$ be the matrices of responses, marginal mean responses, and covariate of the providers in practice i, respectively. The mean of $\boldsymbol{Y}_i$ is denoted by $\boldsymbol{\mu}_i = E(\boldsymbol{Y}_i)$, and the variance of $\boldsymbol{Y}_i$ is $\text{var}(\boldsymbol{Y}_i|\boldsymbol{X}_i) = \theta \boldsymbol{A}_i^{1/2} \boldsymbol{R}_{i0}(\boldsymbol{\omega_0}) \boldsymbol{A}_i^{1/2}$, where $\boldsymbol{A}_i = \text{diag}\{\gamma(\mu_{i11}), \dots, \gamma(\mu_{i1K}), \dots, \gamma(\mu_{in_i1}), \dots, \gamma(\mu_{in_iK})\}$, and a $Kn_i \times Kn_i$ correlation matrix $\boldsymbol{R}_{i0}(\boldsymbol{\omega_0})$ describes the correlation of measures within the ith practice with a vector of association parameters denoted by $\boldsymbol{\omega_0}$. Both γ and θ are dependent on the distribution of responses. If Y_{ijk} is binary, $\gamma(\mu_{ijk}) = \mu_{ijk}(1 - \mu_{ijk})$ and $\theta = 1$. Liang and Zeger[19] showed that $\sqrt{m}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})$ is asymptotically multivariate normal with a covariance matrix

$V_R = \lim_{m\to\infty} m(\Sigma_1^{-1}\Sigma_0\Sigma_1^{-1})$, where $\Sigma_1 = \sum_{i=1}^{m} D_i' V_i^{-1} D_i$, $\Sigma_0 = \sum_{i=1}^{m} D_i' V_i^{-1} \mathrm{cov}(Y_i|X_i) V_i^{-1} D_i$, $D_i = \partial \mu_i / \partial \beta'$, and V_i is a working covariance matrix of Y_i. Let $R_{iw}(\omega)$ be a $Kn_i \times Kn_i$ working correlation matrix with a vector of association parameters ω. The working covariance matrix is expressed as $V_i = \theta A_i^{1/2} R_{iw}(\omega) A_i^{1/2}$ and is unequal to var $(Y_i|X_i)$ unless $R_{iw}(\omega) = R_{i0}(\omega_0)$.

For three-level data, Teerenstra et al proposed a "nested exchangeable" correlation structure[25]:

1. correlation among participants within the same provider in the same practice, is constant, $\mathrm{Corr}(Y_{ijk_1}, Y_{ijk_2}) = r$ for $k_1 \neq k_2$;
2. correlation among participants with different providers in the same practice, is constant, $\mathrm{Corr}(Y_{ij_1k_1}, Y_{ij_2k_2}) = \rho$ for $j_1 \neq j_2$ and any k_1, k_2.

This three-level exchangeable working correlation structure was defined as

$$R_{iw}(r, \rho) = \rho \mathbf{1}_{Kn_i \times Kn_i} + (r - \rho)\,\mathrm{Bdiag}_{n_i}(\mathbf{1}_{K\times K}) + (1 - r)\, I_{Kn_i \times Kn_i},$$

where $\mathbf{1}_{i\times i}$ is an $i \times i$ matrix of 1's, $\mathrm{Bdiag}_i(A)$ is a block diagonal matrix with matrix element A replicated i times, and $I_{i\times i}$ is the $i \times i$ identity matrix. Here, $R_{iw}(r, \rho)$ must be positive definite (PD). Given a value of K and n_i, PD can be determined if the constraints hold: $\min(\lambda_1, \lambda_2, \lambda_{3i}) > 0$, where $\lambda_1 = 1 - r$, $\lambda_2 = 1 + (K-1)r - K\rho$, $\lambda_{3i} = 1 + (K-1)r + K(n_i - 1)\rho$ are the distinct eigenvalues of $R_{iw}(r, \rho)$. The proof was provided in Web Appendix A in the work of Li et al.[26] Here, the constraints are equivalent to

$$-\frac{1}{K-1} < r < 1, -\frac{1+(K-1)r}{K(n_i-1)} < \rho < \frac{1+(K-1)r}{K}. \tag{1}$$

We assume this "nested exchangeable" correlation structure in the following sections.

Suppose we are interested in testing the treatment effect for a two-group comparison: treated versus control. The treatment assignment is coded in the last column of the practice covariate matrices X_i', and the corresponding last parameter of β is β_p. Let V_β denote the (p, p)th element of V_R. Thus, $\sqrt{m}(\widehat{\beta_p} - \beta_p)$ has an asymptotically normal distribution $N(0, V_\beta)$ or, equivalently, $\mathrm{Var}(\widehat{\beta_p}) = V_\beta / m$. For simplicity, we take $p = 2$, ie, coefficient β_1 is the intercept, and coefficient β_2 is the treatment effect. The practice allocations of the treatment and control groups are $m_{\mathrm{trt}} = m\pi$ and $m_{\mathrm{cont}} = m(1 - \pi)$, respectively, where π is a predetermined value, eg, 50%. The hypotheses of interest are $\mathrm{H}_0 : \beta_2 = 0$ versus $\mathrm{H}_1 : \beta_2 = \beta$. For a binary outcome, let p_0 and p_1 be the success rates in the control and treated groups. When the identity link function, $g(\mu_{ijk}) = \mu_{ijk}$, is specified, $\beta_2 = p_1 - p_0$ is the RD between two groups; when the log link function, $g(\mu_{ijk}) = \ln(\mu_{ijk})$, is specified, $\beta_2 = \ln(p_1/p_0)$ is the difference between the natural logarithms of the proportions; and when the logit link function, $g(\mu_{ijk}) = \ln(\mu_{ijk}/(1 - \mu_{ijk}))$, is specified, $\beta_2 = \ln\left(\frac{p_1/(1-p_1)}{p_0/(1-p_0)}\right)$ is the difference between the natural logarithms of the odds. When the log and logit link functions are used, taking the exponential of β_2 refers to the RR and the OR, respectively.

Given the "nested exchangeable" correlation structure, we use identity link function and have

$$\mathrm{Var}(\widehat{\beta_2}) = \frac{\lambda_3}{Knm}\left(\frac{p_1(1-p_1)}{\pi} + \frac{p_0(1-p_0)}{1-\pi}\right), \tag{2}$$

where $n_i \equiv n$ and the eigenvalue $\lambda_3 = 1 + (K-1)r + K(n-1)\rho$. If we consider the log link function, then

$$\mathrm{Var}(\widehat{\beta_2}) = \frac{\lambda_3}{Knm}\left(\frac{1-p_1}{\pi p_1} + \frac{1-p_0}{(1-\pi)p_0}\right). \tag{3}$$

Using the logit link function in the GEE model for a binary outcome, we have

$$\mathrm{Var}(\widehat{\beta_2}) = \frac{\lambda_3}{Knm}\left(\frac{1}{\pi p_1(1-p_1)} + \frac{1}{(1-\pi)p_0(1-p_0)}\right). \tag{4}$$

Please note that Equation (4) is the same as the formula in section 4.4 in the work of Teerenstra et al[25] and reduces to Equation (8)[27] when $K = 1$.

From the relationship between β_2 and RD, the asymptotic variance of $\widehat{RD}$ is

$$\mathrm{Var}(\widehat{RD}) = \frac{\lambda_3}{Knm}\left(\frac{p_1(1-p_1)}{\pi} + \frac{p_0(1-p_0)}{1-\pi}\right). \tag{5}$$

Applying the delta method, we obtain the asymptotic variances of $\widehat{RR}$ and $\widehat{OR}$ as

$$\mathrm{Var}(\widehat{RR}) = \frac{\lambda_3}{Knm}\left(\frac{1-p_1}{\pi p_1} + \frac{1-p_0}{p_0(1-\pi)}\right)\exp\left(\frac{2p_1}{p_0}\right) \tag{6}$$

and

$$\mathrm{Var}(\widehat{OR}) = \frac{\lambda_3}{Knm}\left(\frac{1}{\pi p_1(1-p_1)} + \frac{1}{(1-\pi)\,p_0(1-p_0)}\right)\exp\left(\frac{2p_1/(1-p_1)}{p_0/(1-p_0)}\right). \tag{7}$$

3 | LOCALLY OPTIMAL DESIGN

Assume the study cost per practice is c currency units (eg, US dollars), each provider costs s currency units, and e denotes each participant's cost. The total budget B in a three-level trial is defined as

$$B = m(c + sn + eKn). \tag{8}$$

We aim to find the OD given the constraint in Equation (8). The term "optimal" refers to the variance of each measure estimator being minimized for a given sampling budget.[10,17,28,29]

First, we assume that provider size K is a predetermined value, similarly as B, c, s, and e, for simplicity. The goal is to find the pair of m and n that minimizes the variance of each measure, which is equivalent to maximizing

$$L = \frac{Knm}{\lambda_3} \tag{9}$$

for all three measures (RD, RR, and OR). Substituting $m = \frac{B}{c+(s+eK)n}$ gives

$$L = \frac{KnB}{\lambda_2 c + (\lambda_2 b + K\rho c)\,n + Kb\rho n^2},$$

where $b = s + eK$. Taking the partial derivatives with respect to n gives

$$\frac{\partial L}{\partial n} \propto \lambda_2 c - Kb\rho n^2.$$

Since $\boldsymbol{R}_{iw}(r, \rho)$ is PD, λ_2 is positive. It can be shown that when

$$n = \sqrt{\frac{\lambda_2 c}{Kb\rho}}, \tag{10}$$

where ρ should be positive, the derivatives are equal to 0, and L is maximized. The LOD is reached for a known pair value (r, ρ), and n in Equation (9) is denoted by n_{LOD}. Let $\vartheta = \frac{\lambda_2}{K\rho}$, the parameters in LOD are then given by

$$n_{\mathrm{LOD}} = \sqrt{\frac{\vartheta c}{b}}, \qquad m_{\mathrm{LOD}} = \frac{B}{\sqrt{\vartheta bc} + c}. \tag{11}$$

Please note that $\rho < \frac{[1+(K-1)r]c}{K(c+s+eK)}$ in order to be $n_{\mathrm{LOD}} > 1$. Thus, $0 < \rho < \min\left(\frac{1+(K-1)r}{K}, \frac{[1+(K-1)r]}{K}\frac{c}{(c+s+eK)}\right) = \frac{[1+(K-1)r]c}{K(c+s+eK)}$ since Equation (1) also holds. For any measures (RD, RR, or OR), the LOD is the same even if the variance of the measure estimator is different. Obviously, n_{LOD} and m_{LOD} may be a noninteger. In reality, we need to choose an integer value for practice size with either $n_{\mathrm{up}} =$ int (n_{LOD})+1 or $n_{\mathrm{down}} =$ int (n_{LOD}), where "int" refers to an integer part of a number. We then calculate m_{up} and m_{down} from $m = \frac{B}{c+(s+eK)n}$. Similarly, m_{up} and m_{down} are most likely nonintegers. In order to meet the budget limit, the integer parts for m_{up} and m_{down} are taken as values of the corresponding number of practices. Then, we can calculate the corresponding L using Equation (9), and the proposed optimal practice size and number of practices is the one with the larger L.

Second, when the provider size K is not a fixed value but within a range $(K_{\min}, K_{\max})$ and $K_{\min} \geq 2$, we find n_{LOD} and m_{LOD} for each value of K within this range and calculate the corresponding L in Equation (9). The design with the maximum of L within a range $(K_{\min}, K_{\max})$ is defined as the LOD. Given

$$n_{\mathrm{LOD}} m_{\mathrm{LOD}} = \frac{B}{\sqrt{b}\left(\sqrt{b} + \sqrt{\frac{c}{\vartheta}}\right)},$$

it is easy to show that both $Kn_{\mathrm{LOD}}m_{\mathrm{LOD}}$ and λ_3 are increasing functions of K, but $Kn_{\mathrm{LOD}}m_{\mathrm{LOD}} \propto \sqrt{K}$ and $\lambda_3 \propto K$ when $K \geq 3$. That is, L decreases when K increases for $K \geq 3$. Therefore, the LOD is reached at $K = K_{\min}$ if $K_{\min} \geq 3$ and $K = 3$ if $K_{\min} = 2$ for a known pair value (r, ρ).

Table 1 shows an example to determine the LOD for $r = 0.6$ and $\rho = 0.03$, where $3 \leq K \leq 10$, $B = 300\,000$, $c = 10\,000$, $s = 100$, and $e = 10$ are assumed. For each K, n and m are calculated from Equation (11). Both integers are chosen as

TABLE 1 Locally optimal design for $B = 300\,000$, $c = 10\,000$, $s = 100$, and $e = 10$ with known correlations $r = 0.6$ and $\rho = 0.03$

K	Practice Size n	Number of Practices m	L	Flag	Power[a]	Power[b]	Power[c]
3	43	18	388.3	1	0.871	0.850	0.859
4	40	18	385.0		0.868	0.847	0.856
5	39	18	385.7		0.869	0.848	0.857
6	37	18	381.3		0.865	0.844	0.853
7	36	18	379.6		0.863	0.842	0.851
8	34	18	373.2		0.858	0.836	0.845
9	33	18	370.2		0.855	0.833	0.843
10	32	18	366.9		0.852	0.830	0.839

n and m are calculated from Equation (11).
L is calculated from Equation (9).
Flag = 1 refers to locally optimal design.
[a]Risk difference for $p_0 = 0.3$ and $p_1 = 0.45$.
[b]Risk ratio for $p_0 = 0.3$ and $p_1 = 0.45$.
[c]Odds ratio for $p_0 = 0.3$ and $p_1 = 0.45$.

discussed previously, and the corresponding L is calculated from Equation (9). The design with $K = 3$, $n = 43$, and $m = 18$ is the LOD since L is maximized at $K = 3$. Please note L is not monotonically decreasing in Table 1 since the calculations are provided for (n, m) as integers only. The power estimates for RD, RR, and OR are provided for $p_0 = 0.3$ and $p_1 = 0.45$. It definitely demonstrates that the power is maximized, equivalently minimizing the variance, when the LOD is reached.

4 | MAXIMIN OPTIMAL DESIGN

First, we still assume that the provider size K is a predetermined value. Obviously, n_{LOD} in Equation (11) depends on (r, ρ). In practice, the pair value of (r, ρ) could be unknown before a study starts. If the ranges, $(r_{\min}, r_{\max})$ and $(\rho_{\min}, \rho_{\max})$, can be obtained from previous studies or other literature, then we define them as the parameter space.[30,31] The range of practice size based on practical feasibility, $(n_{\min}, n_{\max})$, is defined as the design space.[17,32,33] The objective is to identify the OD within the parameter and design spaces.

Inserting (11) in (5)-(7) gives the variance of each measure estimator for the OD. For example,

$$\text{Var}(\widehat{RD}) = g(r, \rho) \times \frac{1}{B}\left(\frac{p_1(1-p_1)}{\pi} + \frac{p_0(1-p_0)}{1-\pi}\right), \tag{12}$$

where $g(r, \rho) = \left(\sqrt{\rho c} + \sqrt{\frac{1+(K-1)r-K\rho}{K}(s+eK)}\right)^2$.

Following the same definition of RE,[17] the ratio of the variance of each measure estimator for practice size n_{LOD} to n, we use Equations (5), (8), and (12) and then define RE for measure RD as a function of n, r, and ρ, ie,

$$\text{RE}(n, r, \rho) = \frac{g(r, \rho)}{1 + (K-1)\,r + K(n-1)\rho} \times \frac{Kn}{c + (s + eK)\,n}. \tag{13}$$

It is easy to show that REs for both the RR and OR measures are the same as Equation (13). Furthermore, the maximal value of $\text{RE}(n, r, \rho)$ is 1 and reached when n is n_{LOD}. Figure 1 shows how RE changes across practice size n for a fixed r, and Figure 2 shows the trend of RE over practice size n for a fixed ρ, where $K = 3$, $B = 300\,000$, $c = 10\,000$, $s = 100$, and $e = 10$. Both Figures demonstrate that RE increases until it reaches 1 and then decreases as practice size n increases. Among the four REs with the different values of ρ in Figure 1, we observe that the practice size n at which RE is equal to 1 is the smallest when $\rho = 0.7$ and the largest when $\rho = 0.1$. Similarly, we notice that the practice size n at which RE is equal to 1 is the smallest when $r = 0.1$ and the largest when $r = 0.7$ in Figure 2.

MMD is a design that maximizes some measure of performance (or minimize the risk) in the worst case scenario.[31-34] Here, we use RE, quantified as Equation (13), as the measure of performance. Specifically, the MMD includes three steps. Step 1 defines the parameter and design spaces, Step 2 computes the LOD for each pair value of (r, ρ) in the parameter space and then computes the RE of each design in the design space, and Step 3 finds its smallest RE value within the

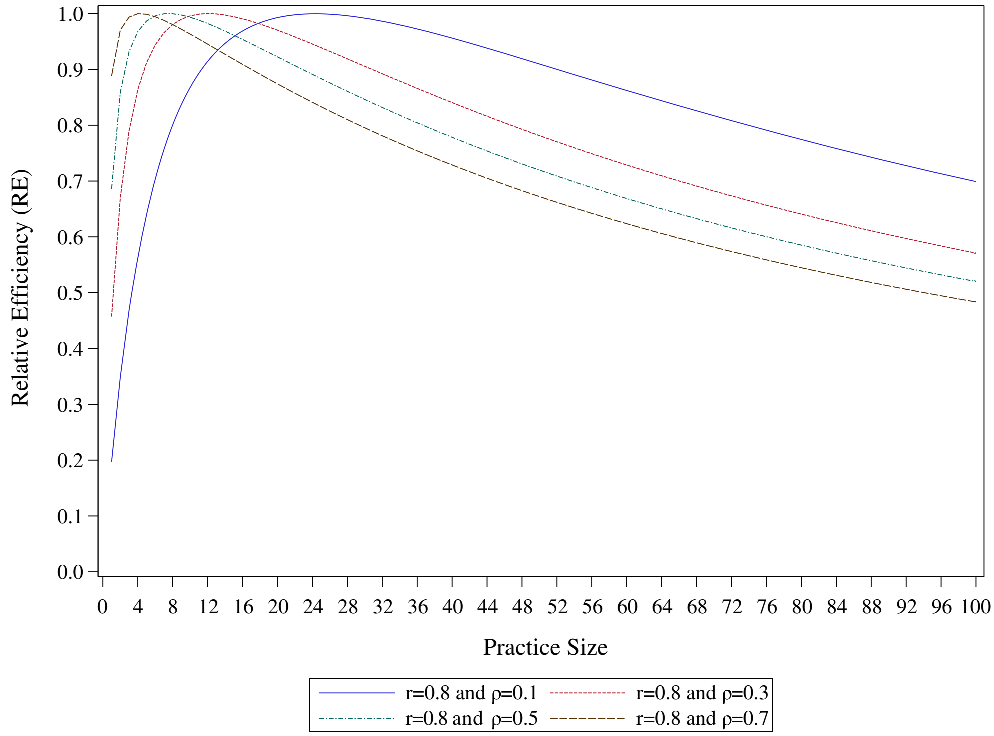

FIGURE 1 Relative efficiencies RE(n, r, ρ) as a function of n for $K = 3$, $B = 300\,000$, $c = 10\,000$, $s = 100$, and $e = 10$ with $r = 0.8$ [Colour figure can be viewed at wileyonlinelibrary.com]

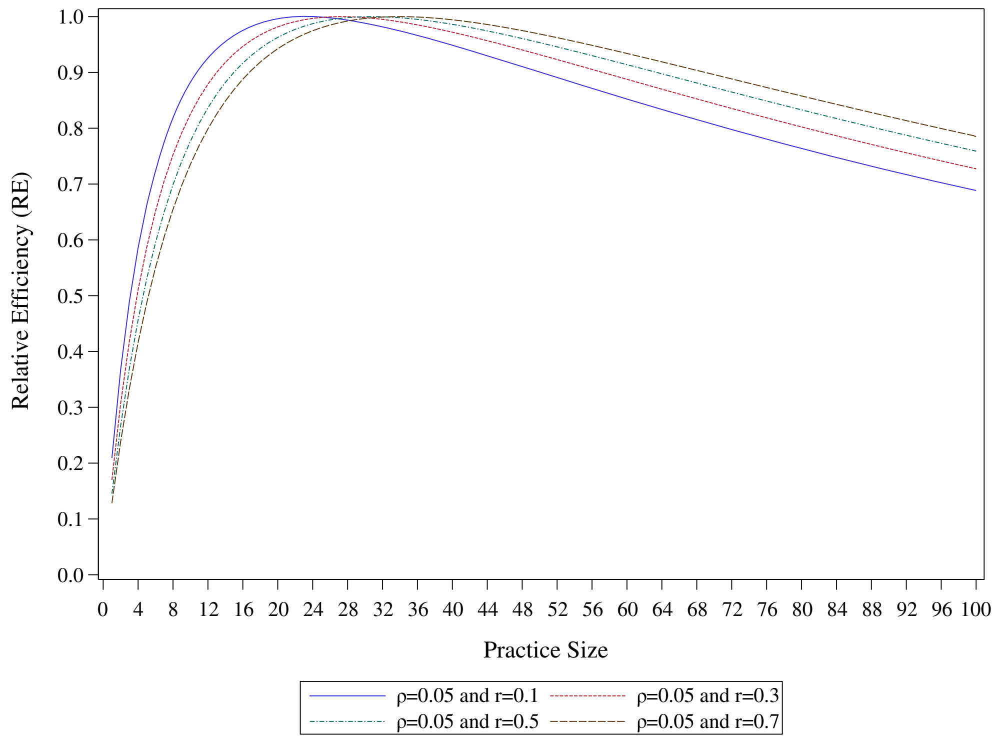

FIGURE 2 Relative efficiencies RE(n, r, ρ) as a function of n for $K = 3$, $B = 300\,000$, $c = 10\,000$, $s = 100$, and $e = 10$ with $\rho = 0.05$ [Colour figure can be viewed at wileyonlinelibrary.com]

parameter space for each design in the design space and selects the design that maximizes the minimum RE among all designs in the design space. This MMD considers the worst case scenario and, thus, is robust against misspecification of the values of (r, ρ).

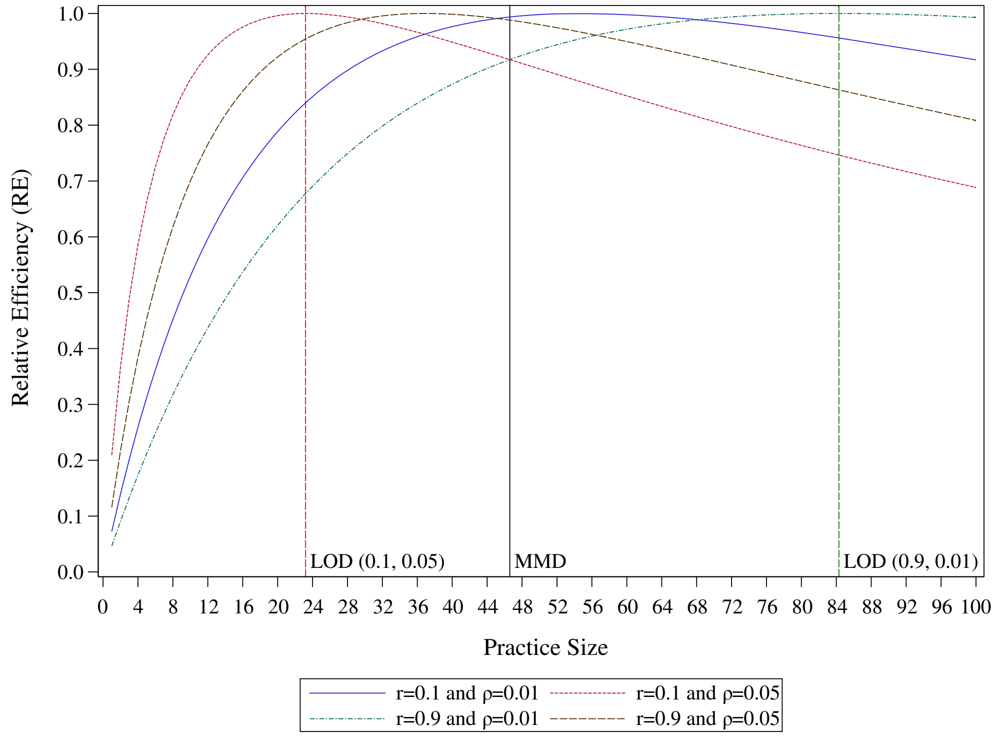

FIGURE 3 Relative efficiencies $\mathrm{RE}(n, r_{\min}, \rho_{\min})$, $\mathrm{RE}(n, r_{\min}, \rho_{\max})$, $\mathrm{RE}(n, r_{\max}, \rho_{\min})$, and $\mathrm{RE}(n, r_{\max}, \rho_{\max})$ as a function of n and locally optimal designs $\mathrm{LOD}(r_{\min}, \rho_{\max})$ and $\mathrm{LOD}(r_{\max}, \rho_{\min})$ for $K = 3$, $B = 300\,000$, $c = 10\,000$, $s = 100$, and $e = 10$ with parameter space $(r_{\min} = 0.1, r_{\max} = 0.9)$ and $(\rho_{\min} = 0.01, \rho_{\max} = 0.05)$. MMD, MaxiMin design [Colour figure can be viewed at wileyonlinelibrary.com]

The RE of any of the three measures, shown in Equation (13), is a function of n, r, and ρ given the costs c per practice, s per provider, e per participant, and the provider size K. First, Appendix A.1 proves that $\mathrm{RE}(n, r, \rho)$ is minimized at one of the four points, namely, $(r_{\min}, \rho_{\min})$, $(r_{\min}, \rho_{\max})$, $(r_{\max}, \rho_{\min})$, $(r_{\max}, \rho_{\max})$, ie, the boundary of the parameter space $(r_{\min}, r_{\max})$ and $(\rho_{\min}, \rho_{\max})$. Figure 3 presents $\mathrm{RE}(n, r_{\min}, \rho_{\min})$, $\mathrm{RE}(n, r_{\min}, \rho_{\max})$, $\mathrm{RE}(n, r_{\max}, \rho_{\min})$, and $\mathrm{RE}(n, r_{\max}, \rho_{\max})$ as functions of n for $K = 3$, $B = 300\,000$, $c = 10\,000$, $s = 100$, and $e = 10$ with parameter space $(r_{\min} = 0.1, r_{\max} = 0.9)$ and $(\rho_{\min} = 0.01, \rho_{\max} = 0.05)$. Next, Appendix A.2 shows that the minimum of $\mathrm{RE}(n, r, \rho)$ is maximized by the design satisfying $\mathrm{RE}(n, r_{\min}, \rho_{\max}) = \mathrm{RE}(n, r_{\max}, \rho_{\min})$. Let $\hat{n}$ be a solution of $\mathrm{RE}(n, r_{\min}, \rho_{\max}) = \mathrm{RE}(n, r_{\max}, \rho_{\min})$ and expressed as

$$\frac{[1 + (K-1)\, r_{\max} - K\rho_{\min}]\, g(r_{\min}, \rho_{\max}) - [1 + (K-1)\, r_{\min} - K\rho_{\max}]\, g(r_{\max}, \rho_{\min})}{K(\rho_{\max} g(r_{\max}, \rho_{\min}) - \rho_{\min} g(r_{\min}, \rho_{\max}))}. \tag{14}$$

As shown in Figure 3, the black vertical straight line indicates $\hat{n}$, and LODs for $(r_{\min}, \rho_{\max})$ and $(r_{\max}, \rho_{\min})$ are added as references. By dividing $g(r_{\min}, \rho_{\max})$ and $g(r_{\max}, \rho_{\min})$ by the study cost per practice c, we notice that MMD of practice sizes depends on $(r_{\min}, \rho_{\max})$, $(r_{\max}, \rho_{\min})$, and ratio $(s + eK)/c$. That is, the total budget B determines the number of practices m but not practice size n.

Now, we provide a step-by-step approach to find an MMD for a two-arm three-level CRT with a binary outcome when the provider size K is a predetermined value.

Step 1: Define the parameter space $(r_{\min}, r_{\max})$, $(\rho_{\min}, \rho_{\max})$ and design space $(n_{\min}, n_{\max})$, respectively.

Step 2: Calculate $\hat{n}$ using Equation (14).

(A) If it is within the range $(n_{\min}, n_{\max})$, then set $n_{\mathrm{MMD}} = \hat{n}$ and the corresponding $m_{\mathrm{MMD}} = \mathrm{int}\left(\frac{B}{c+(s+eK)n_{\mathrm{MMD}}}\right)$.

(B) If it is outside of $(n_{\min}, n_{\max})$, calculate $\mathrm{RE}(n, r_{\min}, \rho_{\min})$, $\mathrm{RE}(n, r_{\min}, \rho_{\max})$, $\mathrm{RE}(n, r_{\max}, \rho_{\min})$, and $\mathrm{RE}(n, r_{\max}, \rho_{\max})$ for each practice size $n \in (n_{\min}, n_{\max})$ and take their minimum. Choose the design of (n,m) that has the maximum of minimum RE within the design space, where $m = \mathrm{int}\left(\frac{B}{c+(s+eK)n}\right)$.

Again, $\hat{n}$ may be a noninteger. We use the same method in Section 3 to get the integer practice size and number of practices. Please note that Equation (13) is derived using Equation (8) as well. If the calculated m_{MMD} from the above approach is infeasible, then the range $(n_{\min}, n_{\max})$ needs to be revised appropriately.

TABLE 2 MaxiMin design for $K = 3$, $B = 300\,000$, $c = 10\,000$, $s = 100$, and $e = 10$ with parameter space ($r_{\min} = 0.1, r_{\max} = 0.9$) and ($\rho_{\min} = 0.01, \rho_{\max} = 0.05$)

Design Space $(n_{\min}, n_{\max})$	Practice Size n	$RE(n, r_{\min}, \rho_{\min})$	$RE(n, r_{\min}, \rho_{\max})$	$RE(n, r_{\max}, \rho_{\min})$	$RE(n, r_{\max}, \rho_{\max})$	Min RE	Number of Practices m	Flag
(11, 20)	11	0.5642	0.9059	0.4090	0.7346	0.4090	26	
	12	0.5966	0.9257	0.4369	0.8656	0.4369	25	
	13	0.6288	0.9421	0.4636	0.7935	0.4636	25	
	14	0.6550	0.9556	0.4892	0.8184	0.4892	25	
	15	0.6813	0.9667	0.5136	0.8408	0.5136	25	
	16	0.7059	0.9757	0.5369	0.8609	0.5369	24	
	17	0.7287	0.9829	0.5592	0.8788	0.5592	24	
	18	0.7501	0.9886	0.5806	0.8949	0.5806	24	
	19	0.7700	0.9929	0.6010	0.9093	0.6010	24	
	20	0.7886	0.9961	0.6205	0.9221	0.6205	23	1
(41, 50)	41	0.9799	0.9441	0.8809	0.9975	0.8809	19	
	42	0.9831	0.9394	0.8881	0.9963	0.8881	19	
	43	0.9859	0.9347	0.8950	0.9949	0.8950	19	
	44	0.9884	0.9299	0.9016	0.9932	0.9016	19	
	45	0.9907	0.9251	0.9079	0.9913	0.9079	18	
	46	0.9926	0.9202	0.9138	0.9893	0.9138	18	
	47	0.9943	0.9154	0.9195	0.9872	0.9154	18	1
	48	0.9958	0.9105	0.9249	0.9849	0.9105	18	
	49	0.9970	0.9056	0.9301	0.9825	0.9056	18	
	50	0.9980	0.9008	0.9350	0.9799	0.9008	18	

Relative efficiency (RE) is calculated from Equation (13).
Flag = 1 refers to MaxiMin design.

TABLE 3 MaxiMin design for $B = 300\,000$, $c = 10\,000$, $s = 100$, and $e = 10$ with parameter space ($r_{\min} = 0.1, r_{\max} = 0.9$) and ($\rho_{\min} = 0.01, \rho_{\max} = 0.05$)

Design Space $(n_{\min}, n_{\max})$	K	Practice Size n	Number of Practices m	RE	Flag
(11, 20)	3	20	23	0.6205	
	4	20	23	0.6369	
	5	20	23	0.6517	
	6	20	22	0.6653	
	7	20	22	0.6781	
	8	20	22	0.6901	
	9	20	21	0.7014	
	10	20	21	0.7121	1
(41, 50)	3	47	18	0.9154	1
	4	43	18	0.9032	
	5	41	18	0.8876	
	6	41	18	0.8638	
	7	41	17	0.8421	
	8	41	17	0.8222	
	9	41	16	0.8037	
	10	41	16	0.7866	

n and m are calculated from the step-by-step approach.
Relative efficiency (RE) is calculated from Equation (13).
Flag = 1 refers to MaxiMin design.

Table 2 shows an example to determine MMD, where the same setting as Figure 3 is assumed. We obtain $\hat{n} = 46.6$ using Equation (13). If the design space is (11, 20), then the design of ($n = 20$, $m = 23$) is MMD under the budget constraints; on the other hand, if the design space is (41, 50), then the design of ($n = 47$, $m = 18$) is MMD under the budget constraints.

Second, when the provider size K is not a fixed value but within a range $(K_{\min}, K_{\max})$ and $K_{\min} \geq 2$, we find n_{MMD} and m_{MMD} for each value of K within this range and calculate the corresponding RE in Equation (13). The design with the maximum of RE is defined as MMD.

TABLE 4 MaxiMin design for $B = 300\,000$, $c = 10\,000$, $s = 100$, and $e = 10$ with $3 \le K \le 10$

Design Space $(n_{\min}, n_{\max})$	$r_{\min}$	$r_{\max}$	$\rho_{\min}$	$\rho_{\max}$	K	Practice Size n	Number of Practices m	RE
(2, 20)	0.1	0.9	0.01	0.05	10	20	21	0.7121
	0.1	0.3	0.01	0.05	10	20	21	0.8717
	0.3	0.6	0.01	0.05	10	20	21	0.7754
	0.6	0.9	0.01	0.05	10	20	21	0.7121
	0.1	0.9	0.01	0.02	10	20	21	0.7121
	0.1	0.9	0.02	0.03	10	20	21	0.8365
	0.1	0.9	0.02	0.05	10	20	21	0.8365
	0.1	0.9	0.03	0.05	10	20	21	0.9031
(2, 50)	0.1	0.9	0.01	0.05	3	47	18	0.9154
	0.1	0.3	0.01	0.05	3	41	19	0.9441
	0.3	0.6	0.01	0.05	3	47	18	0.9446
	0.6	0.9	0.01	0.05	4	50	17	0.9446
	0.1	0.9	0.01	0.02	5	49	17	0.9466
	0.1	0.9	0.02	0.03	5	43	19	0.9751
	0.1	0.9	0.02	0.05	3	41	19	0.9441
	0.1	0.9	0.03	0.05	3	41	19	0.9441

Abbreviation: RE, relative efficiency.

Table 3 demonstrates how to find MMD with parameter space ($r_{\min} = 0.1, r_{\max} = 0.9$) and ($\rho_{\min} = 0.01, \rho_{\max} = 0.05$), where $3 \le K \le 10$, $B = 300\,000$, $c = 10\,000$, $s = 100$, and $e = 10$ are assumed. For each K, n_{MMD} and m_{MMD} are calculated from the step-by-step approach, and the corresponding RE is provided. If the design space is (11, 20), then the design of ($K = 10, n = 20$, $m = 21$) is MMD under the budget constraints; on the other hand, if the design space is (41, 50), then the design of ($K = 3, n = 47$, $m = 18$) is MMD under the budget constraints. SAS macros %OD_3Level_FixedK and %OD_3Level_RangeK are developed to find LOD and MMD when the corresponding parameters are provided.

Last, we conduct a sensitivity analysis about the parameter space. The following eight different parameter spaces are considered: ($r_{\min} = 0.1, r_{\max} = 0.9$) and ($\rho_{\min} = 0.01, \rho_{\max} = 0.05$), ($r_{\min} = 0.1, r_{\max} = 0.3$) and ($\rho_{\min} = 0.01, \rho_{\max} = 0.05$), ($r_{\min} = 0.3, r_{\max} = 0.6$) and ($\rho_{\min} = 0.01, \rho_{\max} = 0.05$), ($r_{\min} = 0.6, r_{\max} = 0.9$) and ($\rho_{\min} = 0.01, \rho_{\max} = 0.05$), ($r_{\min} = 0.1, r_{\max} = 0.9$) and ($\rho_{\min} = 0.01, \rho_{\max} = 0.02$), ($r_{\min} = 0.1, r_{\max} = 0.9$) and ($\rho_{\min} = 0.02, \rho_{\max} = 0.03$), ($r_{\min} = 0.1, r_{\max} = 0.9$) and ($\rho_{\min} = 0.02, \rho_{\max} = 0.05$), and ($r_{\min} = 0.1, r_{\max} = 0.9$) and ($\rho_{\min} = 0.03, \rho_{\max} = 0.05$). We still assume $3 \le K \le 10$, $B = 300\,000$, $c = 10\,000$, $s = 100$, and $e = 10$. Table 4 shows the MMDs for two different design spaces (2, 20) and (2, 50). If $n_{\max}$ is relatively small, eg, $< \hat{n}$, then MMDs are the same ($K = 10, n = 20$, $m = 21$). They might be different otherwise. That is, MMDs are insensitive to the parameter space when the maximum of practice size is relatively small.

5 | EXAMPLE

Teerenstra et al discussed the Helping Hands trial (Netherlands Organization for Health Research and Development ZonMw, grant number 80–007028–98-07101).[25] This study aimed to change nurse behavior through two strategies and randomized the wards to either strategy. The two strategies included the state-of-the-art strategy, which is derived from literature regarding education, reminders, feedback, and targeting adequate products and facilities, and the extended strategy, which contains all elements of the state-of-the-art strategy plus activities aimed at influencing social influence in groups and enhancing leadership. The primary endpoint was adherence to hygiene guidelines (Yes vs. No), and multiple evaluations of nurses' guideline adherence were observed. The researchers expected to improve the adherence from 60% in the state-of-the-art strategy to 70% in the extended strategy. Teerenstra et al considered the constant behavior of nurse $r = 0.6$ and intraward coefficient correlation $\rho = 0.03$.[25] We calculated the total number of wards $m = 58$ to obtain 80% power using the number of nurses per ward $n = 15$ and the number of evaluations $K = 3$ under the same assumptions of (r, ρ) using Equation (4). We assume $c = 2000$, $s = 50$, and $e = 10$ in this study, then the total cost of $58 \times (2000 + 50 \times 15 + 10 \times 3 \times 15) = 185\,600$ will be needed.

We now apply LOD and MMD approaches to redesign this study with the same budget of $B = 185\,600$. We consider $3 \le K \le 6$ and find that LOD is $K = 3, n = 25$, and $m = 46$. The power is 83.7% under this scenario. It is worth mentioning that our proposed method does not guarantee obtaining the desired power, eg, 80%, but to have the highest power under the budget constraints. Researchers should increase the budget if our proposed method does not reach the desired power.

On the other hand, if the researchers have no clear picture of these two associations, then the parameter space need to be specified. Campbell et al showed the ICC interquartile range of implantation studies in secondary care from 0.017 to 0.221.[3] As Teerenstra et al mentioned, the behavior of an individual nurse with respect to hand hygiene is constant; the parameter space $0.5 \leq r \leq 0.9$ is reasonably assumed. Now, the parameter space lies within ($r_{\min} = 0.5, r_{\max} = 0.9$) and ($\rho_{\min} = 0.017, \rho_{\max} = 0.221$), and the design space is set as (3, 50), then the number of evaluations $K = 3$, the number of nurses per ward $n = 17$, and the total number of wards $m = 55$ is our proposed MMD given the budget of $B = 185\,600$.

6 | DISCUSSION

In this paper, we have presented ODs based on GEE models in three-level CRTs and proposed both LODs and MMDs under budget constraints. We employed a nested exchangeable correlation structure[25] and derived the variance of the treatment effect under the assumption of an equal practice size and the same provider size. We derived the LOD when the correlation among participants within the same provider in the same practice, r, and correlation among participants with different providers in the same practice, ρ, are known; the LOD aims to minimize the variance of each measure estimator for a given sampling budget. If the correlation pair (r, ρ) is unknown but lies in a known range, we proposed MMDs for three-level CRTs for a range of r and ρ. We also developed SAS macros to find the LOD and MMD for practical use.

Our method can be extended in several directions. First, our proposed approach is based on the nested exchangeable correlation structure only. It is suitable when the lowest-level units are exchangeable within the middle-level units ("providers") and the middle-level units are exchangeable within the highest-level units ("practices").[25] We will consider more sophisticated settings, eg, a longitudinal data setting with AR(1) correlation among repeated measures over time in our future work. Second, we assume the same practice size and the same provider size. If the practice size is different across providers, the variances of the estimator of treatment effects are more complicated than Equations (2)-(4). The derivation of these formulas warrants further research. Third, when GEE models with the identity or log link are used to analyze correlated binary data, convergence issues may occur since the predicted probability is unconstrained. Fourth, the empirical sandwich estimator of the covariance matrix obtained from GEE is biased for a small number of clusters and, thus, can inflate type I error rates. The proposed LOD and MMD based on the asymptotic variance might be worthy of further investigation. Finally, it merits further consideration to extend treatment groups to more than two.

ACKNOWLEDGEMENTS

We thank the Alvin J. Siteman Cancer Center at Washington University School of Medicine and Barnes-Jewish Hospital in St. Louis, Missouri, for supporting this research (P30 CA91842). The work of Lei Liu was supported by Washington University Institute of Clinical and Translational Sciences grant UL1TR000448 from the National Center for Advancing Translational Sciences of the National Institutes of Health (NIH). The content is solely the responsibility of the authors and does not necessarily represent the official view of the NIH.

CONFLICT OF INTEREST

The authors have declared no conflict of interest.

ORCID

Jingxia Liu https://orcid.org/0000-0002-9434-7907
Lei Liu https://orcid.org/0000-0003-1844-338X
Graham A. Colditz https://orcid.org/0000-0002-7307-0291

REFERENCES

1. Brownson RC, Colditz GA, Proctor EK. *Dissemination and Implementation Research in Health: Translating Science to Practice*. Oxford, UK: Oxford University Press; 2018.
2. James AS, Richardson V, Wang JS, Proctor EK, Colditz GA. Systems intervention to promote colon cancer screening in safety net settings: protocol for a community-based participatory randomized controlled trial. *Implementation Science*. 2013;8:58.

3. Campbell MK, Mollison J, Steen N, Grimshaw JM, Eccles M. Analysis of cluster randomized trials in primary care: a practical approach. *Family Practice*. 2000;17(2):192-196.
4. Gulliford MC, van Staa TP, McDermott L, McCann G, Charlton J, Dregan A. Cluster randomized trials utilizing primary care electronic health records: methodological issues in design, conduct, and analysis (eCRT study). *Trials*. 2014;15:220.
5. Gravenstein S, Dahal R, Gozalo PL, et al. A cluster randomized controlled trial comparing relative effectiveness of two licensed influenza vaccines in US nursing homes: design and rationale. *Clinical Trials*. 2016;13(3):264-274.
6. Kalfon P, Mimoz O, Loundou A, et al. Reduction of self-perceived discomforts in critically ill patients in French intensive care units: study protocol for a cluster-randomized controlled trial. *Trials*. 2016;17(1):87.
7. Mehring M, Haag M, Linde K, Wagenpfeil S, Schneider A. Effects of a web-based intervention for stress reduction in primary care: a cluster randomized controlled trial. *J Med Internet Res*. 2016;18(2):e27.
8. Nagayama H, Tomori K, Ohno K, et al. Effectiveness and cost-effectiveness of occupation-based occupational therapy using the aid for decision making in occupation choice (ADOC) for older residents: pilot cluster randomized controlled trial. *PLoS ONE*. 2016;11(3):e0150374.
9. Yamagata K, Makino H, Iseki K, et al. Effect of behavior modification on outcome in early- to moderate-stage chronic kidney disease: a cluster-randomized trial. *PLoS ONE*. 2016;11(3):e0151422.
10. Raudenbush SW. Statistical analysis and optimal design for cluster randomized trials. *Psychological Methods*. 1997;2(2):173-185.
11. Raudenbush SW, Liu X. Statistical power and optimal design for multisite trials. *Psychological Methods*. 2000;5(2):199-213.
12. Moerbeek M, Van Breukelen GJP, Berger MPF. Optimal experimental design for multilevel logistic models. *The Statistician*. 2001;50(1):17-30.
13. Moerbeek M, Van Breukelen GJP, Berger MPF. Optimal experimental designs for multilevel models with covariates. *Commun Stat Theory Methods*. 2001;30(12):2683-2697.
14. Connelly LB. Balancing the number and size of sites: an economic approach to the optimal design of cluster samples. *Control Clin Trials*. 2003;24(5):544-559.
15. Headrick TC, Zumbo BD. On optimizing multi-level designs: power under budget constraints. *Aust N Z J Stat*. 2005;47(2):219-229.
16. Liu X. Statistical power and optimum sample allocation ratio for treatment and control having unequal costs per unit of randomization. *J Educ Behav Stat*. 2003;28(3):231-248.
17. Van Breukelen G, Candel M. Efficient design of cluster randomized and multicentre trials with unknown intraclass correlation. *Stat Methods Med Res*. 2015;24(5):540-556.
18. Wu S, Wong WK, Crespi CM. Maximin optimal designs for cluster randomized trials. *Biometrics*. 2017;73(3):916-926.
19. Liang K-Y, Zeger SL. Longitudinal data analysis using generalized linear models. *Biometrika*. 1986;73(1):13-22.
20. Toriola AT, Liu J, Ganz PA, et al. Effect of weight loss on bone health in overweight/obese postmenopausal breast cancer survivors. *Breast Cancer Res Treat*. 2015;152(3):637-643.
21. Lin CC, Bruinooge SS, Kirkwood MK, et al. Association between geographic access to cancer care, insurance, and receipt of chemotherapy: geographic distribution of oncologists and travel distance. *J Clin Oncol*. 2015;33(28):3177-3185.
22. Park YH, Jung KH, Im S-A, et al. Quality of life (QoL) in metastatic breast cancer patients with maintenance paclitaxel plus gemcitabine (PG) chemotherapy: results from phase III, multicenter, randomized trial of maintenance chemotherapy versus observation (KCSG-BR07–02). *Breast Cancer Res Treat*. 2015;152(1):77-85.
23. Jeffe DB, Pérez M, Cole EF, Liu Y, Schootman M. The effects of surgery type and chemotherapy on early-stage breast cancer patients' quality of life over 2-year follow-up. *Ann Surg Oncol*. 2016;23(3):735-743.
24. Sanda MG, Dunn RL, Michalski J, et al. Quality of life and satisfaction with outcome among prostate-cancer survivors. *N Engl J Med*. 2008;358(12):1250-1261.
25. Teerenstra S, Lu B, Preisser JS, van Achterberg T, Borm GF. Sample size considerations for GEE analyses of three-level cluster randomized trials. *Biometrics*. 2010;66(4):1230-1237.
26. Li F, Turner EL, Preisser JS. Sample size determination for GEE analyses of stepped wedge cluster randomized trials. *Biometrics*. 2018;74(4):1450-1458.
27. Shih WJ. Sample size and power calculations for periodontal and other studies with clustered samples using the method of generalized estimating equations. *Biometrical Journal*. 1997;39(8):899-908.
28. Moerbeek M, van Breukelen GJP, Berger MPF. Design issues for experiments in multilevel populations. *J Educ Behav Stat*. 2000;25(3):271-284.
29. Liu J, Colditz GA. Optimal design of longitudinal data analysis using generalized estimating equation models. *Biometrical Journal*. 2017;59(2):315-330.
30. Atkinson A, Donev A, Tobias R. *Optimum Experimental Designs, with SAS*. Oxford, UK: Oxford University Press; 2007.
31. Berger MPF, Wong W-K. *An Introduction to Optimal Designs for Social and Biomedical Research*. Chichester, UK: John Wiley & Sons; 2009.
32. Winkens B, Schouten HJA, van Breukelen GJP, Berger MPF. Optimal designs for clinical trials with second-order polynomial treatment effects. *Stat Methods Med Res*. 2007;16(6):523-537.
33. Ouwens MJNM, Tan PES, Berger MPF. Maximin D-optimal designs for longitudinal mixed effects models. *Biometrics*. 2002;58(4):735-741.
34. Maus B, van Breukelen GJ, Goebel R, Berger MP. Robustness of optimal design of fMRI experiments with application of a genetic algorithm. *NeuroImage*. 2010;49(3):2433-2443.

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of the article.

How to cite this article: Liu J, Liu L, Colditz GA. Optimal designs in three-level cluster randomized trials with a binary outcome. *Statistics in Medicine*. 2019;38:3733–3746. https://doi.org/10.1002/sim.8153

APPENDIX

The proof consists of two steps. Appendix A.1 shows that $\mathrm{RE}(n, r, \rho)$ is minimized at one of the four points, namely, $(r_{\min}, \rho_{\min})$, $(r_{\min}, \rho_{\max})$, $(r_{\max}, \rho_{\min})$, $(r_{\max}, \rho_{\max})$, ie, the boundary of the parameter space $(r_{\min}, r_{\max})$ and $(\rho_{\min}, \rho_{\max})$. Next, Appendix A.2 shows that the minimum of $\mathrm{RE}(n, r, \rho)$ is maximized by the design satisfying $\mathrm{RE}(n, r_{\min}, \rho_{\max}) = \mathrm{RE}(n, r_{\max}, \rho_{\min})$.

A.1 | Proof that $\mathrm{RE}(n, r, \rho)$ is minimized at one of the four points: $(r_{\min}, \rho_{\min})$, $(r_{\min}, \rho_{\max})$, $(r_{\max}, \rho_{\min})$, $(r_{\max}, \rho_{\max})$ within the parameter space $(r_{\min}, r_{\max})$ and $(\rho_{\min}, \rho_{\max})$

From Equations (5), (8), and (12), it follows that the RE for measure RD as a function of n, r, and ρ for any of the three measures is

$$\mathrm{RE}(n, r, \rho) = \frac{g(r, \rho)}{1 + (K-1)\,r + K\,(n-1)\,\rho} \times \frac{Kn}{c + (s + eK)\,n}.$$

Taking the partial derivative with respect to ρ gives

$$\frac{\partial \mathrm{RE}(n, r, \rho)}{\partial \rho} \propto \left\{ \left[\sqrt{\frac{c}{\rho}} - \sqrt{\frac{K(s+eK)}{1 + (K-1)\,r - K\rho}} \right] [1 + (K-1)\,r + K(n-1)\rho] - \left(\sqrt{\rho c} + \sqrt{\frac{1 + (K-1)\,r - K\rho}{K}}\,(s + eK) \right) K(n-1) \right\}.$$

Setting the right-hand side to zero, we obtain

$$\rho_* = \frac{c[1 + (K-1)\,r]}{K[c + (s+eK)\,n^2]}.$$

Then, taking the partial derivative with respect to r gives

$$\frac{\partial \mathrm{RE}(n, r, \rho)}{\partial r} \propto [1 + (K-1)\,r + K(n-1)\rho] \sqrt{\frac{s+eK}{K\,[1 + (K-1)\,r - K\rho]}} - \left(\sqrt{\rho c} + \sqrt{\frac{1 + (K-1)\,r - K\rho}{K}}\,(s+eK) \right).$$

Similarly, we set the right-hand side to zero and obtain

$$r_* = \frac{K\rho[c + (s+eK)\,n^2] - c}{c(K-1)}.$$

Both are actually the same as n_{LOD} in Equation (10). We can show that $\frac{\partial \mathrm{RE}(n,r,\rho)}{\partial \rho} > 0$ if $\rho_{\min} \le \rho < \rho_*$ and $\frac{\partial \mathrm{RE}(n,r,\rho)}{\partial \rho} < 0$ if $\rho_* < \rho \le \rho_{\max}$; hence, $\mathrm{RE}(n, r, \rho)$ is minimized at either $\rho = \rho_{\min}$ or $\rho = \rho_{\max}$ for a fixed r. If we assume the possible range for ρ is (0.1, 0.7) and $r = 0.8$, Figure A1 demonstrates a three-dimensional RE plot as a function of n and ρ, whereas Figure 1 shows the RE plots for four paired values (r, ρ). As seen in Figure 1, $\mathrm{RE}(n, r, \rho)$ is minimized at $\rho = 0.1$ when $n < 13$ and at $\rho = 0.7$ when $n > 13$. Similarly, $\frac{\partial \mathrm{RE}(n,r,\rho)}{\partial r} > 0$ if $r_{\min} \le r < r_*$ and $\frac{\partial \mathrm{RE}(n,r,\rho)}{\partial r} < 0$ if $r_* < r \le r_{\max}$; hence, $\mathrm{RE}(n, r, \rho)$ is minimized at either $r = r_{\min}$ or $r = r_{\max}$ for a fixed ρ. Figure A2 demonstrates a three-dimensional RE plot as a function of n and r, $r \in (0.1, 0.7)$, for a fixed $\rho = 0.05$. As shown in Figure 2, with $\rho = 0.05$, $\mathrm{RE}(n, r, \rho)$ is minimized at $r = 0.7$ when $n < 28$ and $r = 0.1$ when $n < 28$. When combining these characteristics, we conclude that $\mathrm{RE}(n, r, \rho)$ is minimized at $(r_{\min}, \rho_{\min})$, or $(r_{\min}, \rho_{\max})$, or $(r_{\max}, \rho_{\min})$, or $(r_{\max}, \rho_{\max})$ within the parameter space $(r_{\min}, r_{\max})$ and $(\rho_{\min}, \rho_{\max})$.

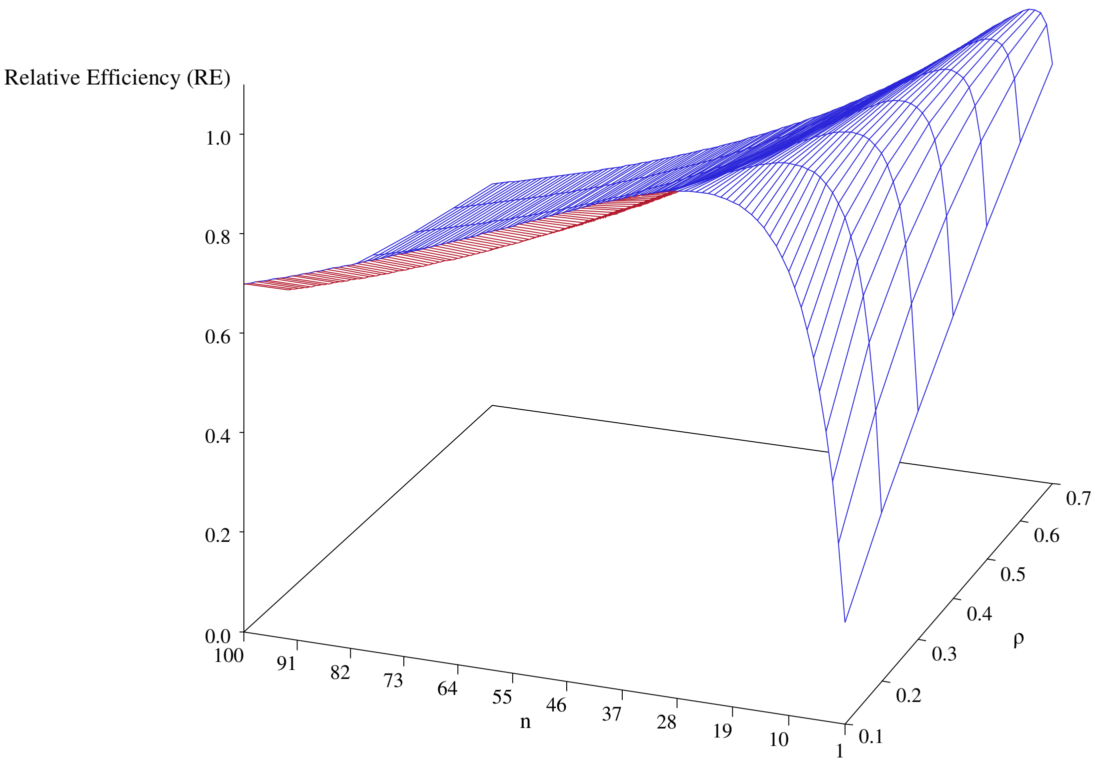

FIGURE A1 Three-dimensional relative efficiencies RE(n, r, ρ) as a function of n and ρ for $K = 3$, $B = 300\,000$, $c = 10\,000$, $s = 100$, and $e = 10$ with $r = 0.8$ [Colour figure can be viewed at wileyonlinelibrary.com]

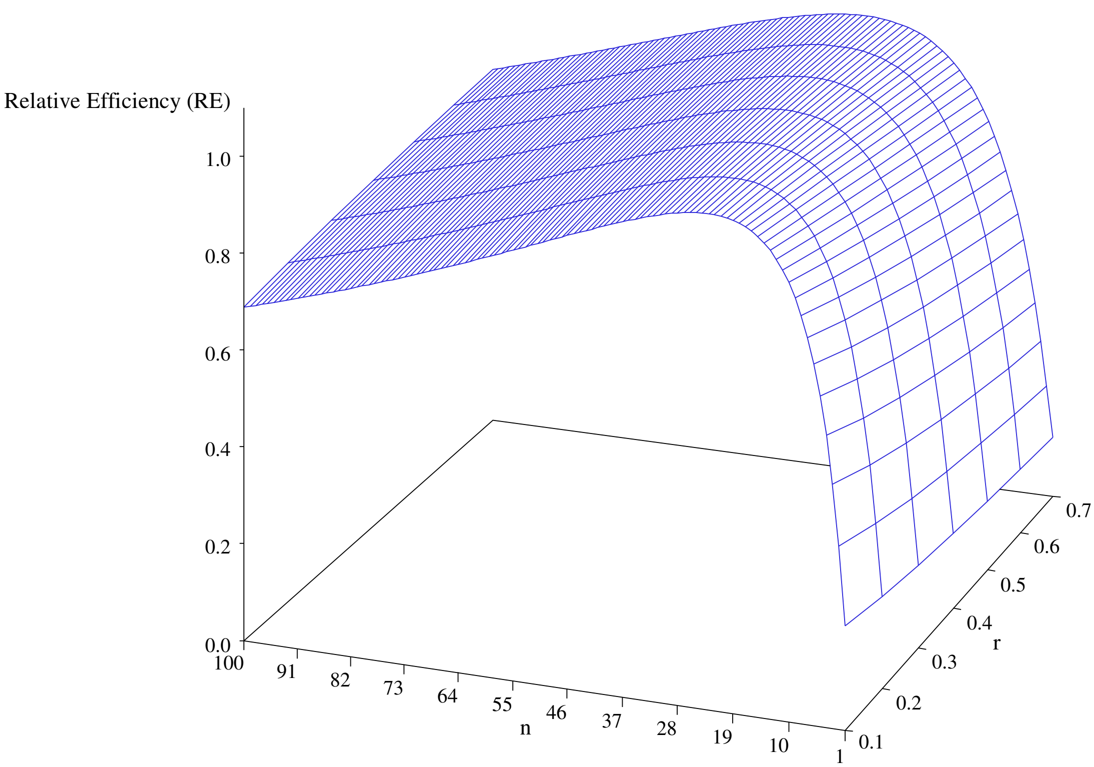

FIGURE A2 Three-dimensional relative efficiencies RE(n, r, ρ) as a function of n and r for $K = 3$, $B = 300\,000$, $c = 10\,000$, $s = 100$, and $e = 10$ with $\rho = 0.05$ [Colour figure can be viewed at wileyonlinelibrary.com]

A.2 | Proof that the minimum of RE(n, r, ρ) is maximized by the design satisfying RE($n, r_{\min}, \rho_{\max}$) = RE($n, r_{\max}, \rho_{\min}$)

Inserting ($r_{\min}, \rho_{\min}$) in Equation (12) and taking the partial derivative with respect to n gives

$$\frac{\partial \mathrm{RE}(n, r_{\min}, \rho_{\min})}{\partial n} \propto ac - Kb\rho n^2.$$

Following the proof in Section 3, RE($n, r_{\min}, \rho_{\min}$) is a single-peaked function and maximized at

$$n_{(r_{\min}, \rho_{\min})} = \sqrt{\frac{1 + (K-1)\, r_{\min} - K\rho_{\min}}{K\rho_{\min}} \times \frac{c}{b}}$$

with a maximum of 1. Similarly, RE($n, r_{\min}, \rho_{\max}$) is maximized at

$$n_{(r_{\min}, \rho_{\max})} = \sqrt{\frac{1 + (K-1)\, r_{\min} - K\rho_{\max}}{K\rho_{\max}} \times \frac{c}{b}},$$

RE($n, r_{\max}, \rho_{\min}$) is maximized at

$$n_{(r_{\max}, \rho_{\min})} = \sqrt{\frac{1 + (K-1)\, r_{\max} - K\rho_{\min}}{K\rho_{\min}} \times \frac{c}{b}},$$

and RE($n, r_{\max}, \rho_{\max}$) is maximized at

$$n_{(r_{\max}, \rho_{\max})} = \sqrt{\frac{1 + (K-1)\, r_{\max} - K\rho_{\max}}{K\rho_{\max}} \times \frac{c}{b}}.$$

Since $\rho_{\min} < \rho_{\max}$, it gives $n_{(r_{\min}, \rho_{\min})} > n_{(r_{\min}, \rho_{\max})}$ and $n_{(r_{\max}, \rho_{\min})} > n_{(r_{\max}, \rho_{\max})}$. Furthermore, $r_{\min} < r_{\max}$ is followed by $n_{(r_{\min}, \rho_{\min})} < n_{(r_{\max}, \rho_{\min})}$ and $n_{(r_{\min}, \rho_{\max})} < n_{(r_{\max}, \rho_{\max})}$. Thus, $n_{(r_{\min}, \rho_{\max})}$ is the smallest, and $n_{(r_{\max}, \rho_{\min})}$ is the largest. All four REs have a maximum of 1, and the maximums are reached at $n_{(r_{\min}, \rho_{\min})}$, $n_{(r_{\min}, \rho_{\max})}$, $n_{(r_{\max}, \rho_{\min})}$, and $n_{(r_{\max}, \rho_{\max})}$, respectively. Following the proof in the appendix of Breukelen et al,[20] the minimum RE between any two of RE($n, r_{\min}, \rho_{\min}$), RE($n, r_{\min}, \rho_{\max}$), RE($n, r_{\max}, \rho_{\min}$), and RE($n, r_{\max}, \rho_{\max}$) is maximized by the design satisfying these two REs are equal. For example, the minimum RE for RE($n, r_{\min}, \rho_{\min}$) and RE($n, r_{\min}, \rho_{\max}$) is maximized by the design satisfying RE($n, r_{\min}, \rho_{\min}$) = ($n, r_{\min}, \rho_{\max}$). For any two pair values, (r_0, ρ_0) and (r_1, ρ_1), the intersection means

$$\mathrm{RE}(n, r_0, \rho_0) = \mathrm{RE}(n, r_1, \rho_1) \xrightarrow{\text{yields}} \frac{g(r_0, \rho_0)}{1 + (K-1)\, r_0 + K\,(n-1)\,\rho_0} = \frac{g(r_1, \rho_1)}{1 + (K-1)\, r_1 + K(n-1)\rho_1}.$$

Its only solution is

$$n = \frac{[1 + (K-1)\, r_1 - K\rho_1]\, g(r_0, \rho_0) - [1 + (K-1)\, r_0 - K\rho_0]\, g(r_1, \rho_1)}{K\,(\rho_0 g(r_1, \rho_1) - \rho_1 g(r_0, \rho_0))}.$$

That is, there is only one intersection between any two RE(n, r, ρ)'s in the function of n. Therefore, there are a total of six intersections across these four REs. For example, Figure 3 demonstrates REs at the four points and all six intersections. Given all facts that these four are single-peaked functions, $n_{(r_{\min}, \rho_{\max})}$ is the smallest and $n_{(r_{\max}, \rho_{\min})}$ is the largest, and the only one intersection between any two REs, it is obvious that the minimum of RE(n, r, ρ) for these six intersections is reached at the intersection of RE($n, r_{\min}, \rho_{\max}$) = RE($n, r_{\max}, \rho_{\min}$). In other words, the minimum of RE(n, r, ρ) is maximized by the design satisfying RE($n, r_{\min}, \rho_{\max}$) = RE($n, r_{\max}, \rho_{\min}$).