HHS Public Access

Author manuscript

Biom J. Author manuscript; available in PMC 2019 September 25.

Published in final edited form as:

Biom J. 2018 May ; 60(3): 616–638. doi:10.1002/bimj.201600262.

Relative efficiency of unequal versus equal cluster sizes in cluster randomized trials using generalized estimating equation models

Jingxia Liu[1], **Graham A. Colditz**[2]

[1]Division of Public Health Sciences, Dept. of Surgery, Washington University in Saint Louis (WUSTL), St Louis, Missouri, USA 63110

[2]Dept. of Surgery, Washington University in Saint Louis (WUSTL), St Louis, Missouri, USA 63110

Abstract

There is growing interest in conducting cluster randomized trials (CRTs). For simplicity in sample size calculation, the cluster sizes are assumed to be identical across all clusters. However, equal cluster sizes are not guaranteed in practice. Therefore, the relative efficiency (RE) of unequal versus equal cluster sizes has been investigated when testing the treatment effect. One of the most important approaches to analyze a set of correlated data is the generalized estimating equation (GEE) proposed by Liang and Zeger, in which the "working correlation structure" is introduced and the association pattern depends on a vector of association parameters denoted by ρ. In this paper, we utilize GEE models to test the treatment effect in a two-group comparison for continuous, binary, or count data in CRTs. The variances of the estimator of the treatment effect are derived for the different types of outcome. RE is defined as the ratio of variance of the estimator of the treatment effect for equal to unequal cluster sizes. We discuss a commonly used structure in CRTs- exchangeable, and derive the simpler formula of RE with continuous, binary, and count outcomes. Finally, REs are investigated for several scenarios of cluster size distributions through simulation studies. We propose an adjusted sample size due to efficiency loss. Additionally, we also propose an optimal sample size estimation based on the GEE models under a fixed budget for known and unknown association parameter (ρ) in the working correlation structure within the cluster.

Keywords

Cluster randomized trial (CRT); generalized estimating equation (GEE); relative efficiency (RE); working correlation structure; intraclass correlation coefficient (ICC)

1 Introduction

In recent years, there has been growing interest in conducting cluster randomized trials (CRTs) (Campbell et al., 2000; Gulliford et al., 2014; Gravenstein et al., 2016; Kalfon et al.,

Conflict of Interest

The authors have declared no conflict of interest.

2016; Mehring et al., 2016; Nagayama et al., 2016; Yamagata et al., 2016). CRTs are performed when randomization at the individual-level is practically infeasible or may lead to severe estimation bias of treatment effect. It is reasonable to expect correlation among individuals in the same cluster, since they may share similar characteristics or be exposed to common factors. The degree of such similarity is commonly quantified by the intracluster correlation coefficient (ICC). Even if the ICCs are pretty small for most CRTs (Murray et al., 2004; Turner et al., 2004), they must be considered in study design and data analysis to avoid underestimation of variance of the treatment effect and inflated type I error rates (Murray, Varnell, et al., 2004).

Methods of analyzing correlated data have been extensively developed. One of the most important approaches is the generalized estimating equation (GEE) proposed by Liang and Zeger (Liang et al., 1986), which fits a marginal model in the context of longitudinal studies. In order to describe the pattern of measures within the individual, the idea of "working correlation structure" is introduced and the pattern depends on a vector of association parameters denoted by ρ. The advantage of this approach is the consistent estimates provided that the marginal model is correctly specified, even if the working correlation matrix is incorrectly assumed. Therefore, GEEs have been commonly applied (Sanda et al., 2008; Lin et al., 2015; Park et al., 2015; Toriola et al., 2015; Jeffe et al., 2016).

Sample size calculation or power estimation is an important topic in study design for investigators. The equations of sample size for CRTs have been published as a function of ICC or the coefficient of variation (CV) of cluster sizes (Donner et al., 1981; Liu et al., 1997; Shih, 1997; Rosner et al., 2003; Eldridge et al., 2006; Austin, 2007; Candel et al., 2008; Van Breukelen et al., 2008; Teerenstra et al., 2010; Rosner et al., 2011; Van Breukelen et al., 2012; Amatya et al., 2013). Shih (Shih, 1997) provides a sample size formula based on the GEE methods for CRTs designed to test the treatment effect. The general formula for a two-group comparison is given. Specifically, the formulas for continuous and binary outcomes are demonstrated with the assumptions of equal cluster sizes n and the exchangeable working correlation R(ρ), i.e., 1 for diagonal entries and ρ otherwise. For count data, Bhaumik and Gibbons (Amatya, Bhaumik, et al., 2013) consider population-average estimators using GEE models for cluster-level randomized design. They assume half of the clusters are randomized to the treatment and the other half are randomized to the control. The asymptotic variance of the treatment effect estimator is derived and the required number of clusters is calculated with the same assumptions: equal cluster sizes n and the exchangeable working correlation R(ρ).

In the design of CRTs, the required sample sizes include the number of clusters m and cluster sizes n_i, $i = 1, \cdots, m$. For simplicity, the cluster sizes are assumed to be identical across clusters, i.e., $n_i = n$ for all i, resulting in total sample size of $N = nm$. In general, the cluster size n is a pre-determined number, therefore the sample size calculation is for the number of clusters only. However, equal cluster sizes are not guaranteed in practice. The relative efficiency (RE) of unequal versus equal cluster sizes has been investigated when testing the treatment effect. For continuous outcomes in cluster randomized studies, Manatunga et al (Manatunga et al., 2001) derive sample size estimation while accounting for the variability due to cluster size and a lower bound for RE is given. For both CRTs and

person randomization within centers, Breukelen et al (Van Breukelen et al., 2007) address the RE of unequal versus equal cluster sizes with continuous outcomes. They investigate RE based on maximum likelihood parameter estimation for a range of cluster size distributions. Additionally, they present an approximate formula for computing the RE as a function of the mean and variance of cluster size and the intraclass correlation (ρ). They conclude that the loss of efficiency due to variation of cluster sizes rarely exceeds 10 percent and can be compensated by sampling 11 percent more clusters. Candel and Breukelen (Candel et al., 2010) give the adjusted sample size formula for varying cluster sizes with a binary outcome in CRTs. They derive the asymptotic RE of unequal versus equal cluster sizes from first-order marginal quasi-likelihood (MQL) estimation with mixed effect logistic regression model. A simpler formula of sample size calculation is presented to estimate the efficiency loss due to variation in cluster sizes. They find that 14% more clusters are needed to cover the efficiency loss in many cases.

In this paper, we will utilize GEE models to test the treatment effect in a two-group comparison for continuous, binary, or count data in CRTs. The variances of the estimator of the treatment effect will be derived for different types of outcome. Obviously, they are dependent on which working correlation structure is chosen. As shown (Shih, 1997; Amatya, Bhaumik, et al., 2013), researchers often consider the exchangeable structure since the independent working correlation structure is a special case of exchangeable and first order auto-regressive (AR(1)) fits time-structured data more appropriately. Here, we will also discuss this commonly used structure. RE is defined as the ratio of variance of the estimator of the treatment effect for equal to unequal cluster sizes. We will derive a simpler formula of RE with continuous, binary, and count outcomes. Finally, REs are investigated for several scenarios of cluster sizes distribution through simulation studies. We will propose the adjusted sample size due to the efficiency loss. Additionally, cost is an important factor in study design of CRTs since it is associated with recruiting an additional cluster instead of an additional subject in an individual-level randomized trial. Under a fixed budget we also have the same proposal about the optimal sample size, which maximizes the power, based on the GEE models for known and unknown association parameter (ρ) in the working correlation structure within the cluster.

The outline of this paper is as follows. In section 2, we briefly summarize the GEE methods proposed by Liang and Zeger (Liang and Zeger, 1986) and derive the variance of the estimator of the treatment for three kinds of outcomes (continuous, binary, and count) in a two-group comparison. Section 3 introduces the REs of unequal versus equal cluster sizes for the treatment effect. The formula for RE is derived for the exchangeable working correlation structure. Section 4 presents the possible patterns of cluster size distribution and the results through simulation studies that investigate the REs. In section 5, we show the sample size formulas under two different situations- no cost constrains and given a fixed budget. The practical uses are discussed when planning a CRT, followed by discussion of the limitations and directions for future research.

2. Generalized estimating equation models

Let $\boldsymbol{y}_i = (y_{i1}, \quad y_{i2}, \cdots, \quad y_{in_i})'$ be a vector of responses from the ith cluster, $i = 1, \cdots, m$. The responses are assumed to be independent across clusters but correlated within each cluster. The marginal model is

$$g\left(\mu_{ij}\right) = \mathbf{C}'_{ij} \varnothing , \qquad i = 1, \quad \cdots, \quad m; \quad \mathrm{j} = 1, \cdots, \mathrm{n}_i, \quad (1)$$

where $\mu_{ij} = E(y_{ij})$ and $g(\cdot)$ is a linking function. $\mathbf{C}_{ij}$ is a vector of covariates, and $\varnothing$ is an unknown $l \times 1$ vector of regression coefficients. This model specifies a relationship between μ_{ij} and covariates $\mathbf{C}_{ij}$. The conditional variance of y_{ij} given $\mathbf{C}_{ij}$ is defined as var $(y_{ij}|\mathbf{C}_{ij}) = \gamma(\mu_{ij})\theta$, where γ is a known variance function of μ_{ij} nd θ is a scale parameter. The mean of $\boldsymbol{y}_i$ is denoted by $\boldsymbol{\mu}_i = E(\boldsymbol{y}_i)$ and the variance-covariance matrix for $\boldsymbol{y}_i$ is denoted by $\boldsymbol{V}_i = \theta \boldsymbol{A}_i^{1/2} \boldsymbol{R}_i(\rho) \boldsymbol{A}_i^{1/2}$, where $\boldsymbol{A}_i = \mathrm{diag}\left\{\gamma\left(\mu_{i1}\right), \cdots, \quad \gamma\left(\mu_{in_i}\right)\right\}$ and working correlation structure $\boldsymbol{R}_i(\rho)$ describes the pattern of measures within the ith cluster. $\boldsymbol{R}_i(\rho)$ is an $n_i \times n_i$ matrix and depends on a vector of association parameters denoted by ρ. Both γ and θ are dependent on the distribution of responses. For instance, if y_{ij} is continuous, $\gamma(\mu_{ij}) = 1$ and θ is the random error variance; if y_{ij} is binary, $\gamma(\mu_{ij}) = \mu_{ij}(1 - \mu_{ij})$ and θ=1; If y_{ij} is count, $\gamma(\mu_{ij}) = \mu_{ij}$ and θ = 1.

The estimate of $\varnothing$ is obtained by solving the following estimating equation

$$U(\varnothing) = \sum_{i=1}^{m} \boldsymbol{D}_i' \mathbf{V}_i^{-1} \left(\boldsymbol{y}_i - \boldsymbol{\mu}_i\right) = 0, \quad (2)$$

Where $\boldsymbol{D}_i = \partial \boldsymbol{\mu}_i / \partial \varnothing'$. Let $\widehat{\varnothing}$ be the solution of Eq. (2). Liang and Zeger (Liang and Zeger, 1986) show that $\sqrt{m}(\widehat{\varnothing} - \varnothing)$ is asymptotically multivariate normal with covariance matrix

$$\boldsymbol{V}_{\boldsymbol{R}} = \lim_{m \to \infty} m\left(\Sigma_1^{-1} \Sigma_0 \Sigma_1^{-1}\right), \quad (3)$$

Where $\Sigma_1 = \sum_{i=1}^{m} \boldsymbol{D}_i' \mathbf{V}_i^{-1} \boldsymbol{D}_i$ and $\Sigma_0 = \sum_{i=1}^{m} \boldsymbol{D}_i' \mathbf{V}_i^{-1} \mathrm{cov}(\boldsymbol{y}_i|\mathbf{C}_i) \mathbf{V}_i^{-1} \boldsymbol{D}_i$. As noted, GEE gives asymptotically consistent estimate $\widehat{\varnothing}$ even when the working correlation matrix is incorrectly assumed (Liang and Zeger, 1986).

For the purpose of sample size calculation at the design stage, we need to assume values for the variance and covariance of $\boldsymbol{y}_i$. Therefore, $\Sigma_0 = \Sigma_1$ and $\boldsymbol{V}_{\boldsymbol{R}} = \lim_{m \to \infty} m\Sigma_1^{-1}$ when cov($\boldsymbol{y}_i$|$\mathbf{C}_i$) is assumed to be same as $\boldsymbol{V}_i$. Suppose we are interested in testing the treatment effect for a two-group comparison: the treated vs. control group. For simplicity, let us consider a special case with $l = 2$, where $\varnothing = (\beta_0, \beta_1)'$ and the design matrices $\mathbf{C}_i$. Specifically, the

coefficient β_0 is the intercept, β_1 is the treatment effect, $\mathbf{C}_{it} = \left(\mathbf{1}_{n_i}, \quad \mathbf{1}_{n_i}\right)$ for ith cluster assigned to the treated group, $\mathbf{C}_{it} = \left(\mathbf{1}_{n_i}, \quad \mathbf{0}_{n_i}\right)$ for ith cluster assigned to the control group, and $\mathbf{1}_n$ and $\mathbf{0}_n$ an $n \times 1$ vector of 1's and 0's, respectively. Let V_β denote the (2, 2)th element of $\boldsymbol{V_R}$. Thus, $\sqrt{m}\left(\widehat{\beta_1} - \beta_1\right)$ has an asymptotically normal distribution $N(0, V_\beta)$, equivalently, $\mathrm{Var}\left(\widehat{\beta_1}\right) = V_\beta/m$. The cluster allocations of the treated and control groups are, $m_t = m\pi$ and $m_c = m(1 - \pi)$, respectively. The hypotheses of interest are $H_0{:}\beta_1 = 0$ versus $H_1{:}\beta_1 = \beta$.

2.1. Continuous outcome

This section shows details of using an identity link function on the continuous outcome, that is, $\boldsymbol{\mu}_i = \mathbf{C}_i' \varnothing$, and $\boldsymbol{V}_i = \sigma^2 \boldsymbol{R}_i(\rho)$. With this setting, $\boldsymbol{D_i} = \partial \boldsymbol{\mu}_i / \partial \varnothing = \mathbf{C}_i$. Thus,

$$\Sigma_1 = \frac{1}{\sigma^2}\sum_{i=1}^{m} \boldsymbol{C}_i' \mathbf{R}_i^{-1} \boldsymbol{C}_i = \frac{1}{\sigma^2}\sum_{i \in \mathrm{trt}} \boldsymbol{C}_{it}' \mathbf{R}_i^{-1} \boldsymbol{C}_{it} + \frac{1}{\sigma^2}\sum_{i \in \mathrm{cont}} \boldsymbol{C}_{it}' \mathbf{R}_i^{-1} \boldsymbol{C}_{it}$$
$$= \frac{1}{\sigma^2}\sum_{i \in \mathrm{trt}} \mathbf{1}_{n_i}' \mathbf{R}_i^{-1} \mathbf{1}_{n_i} \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix} + \frac{1}{\sigma^2}\sum_{i \in \mathrm{cont}} \mathbf{1}_{n_i}' \mathbf{R}_i^{-1} \mathbf{1}_{n_i} \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} = \frac{1}{\sigma^2} \begin{pmatrix} t + c & t \\ t & t \end{pmatrix},$$

Where $i \in$ trt denotes all the clusters assigned to treated group, and $i \in$ cont denotes all the clusters assigned to control group, respectively; $t = \sum_{i \in \mathrm{trt}} \mathbf{1}_{n_i}' \mathbf{R}_i^{-1} \mathbf{1}_{n_i}$ and $c = \sum_{i \in \mathrm{cont}} \mathbf{1}_{n_i}' \mathbf{R}_i^{-1} \mathbf{1}_{n_i}$. Then

$$\boldsymbol{V_R} = \lim_{m \to \infty} m\Sigma_1^{-1} = \sigma^2 \lim_{m \to \infty} m \begin{pmatrix} \frac{1}{c} & -\frac{1}{c} \\ -\frac{1}{c} & \frac{1}{t} + \frac{1}{c} \end{pmatrix}, \text{ and}$$

$$\boldsymbol{V}_\beta = \sigma^2 \lim_{m \to \infty} m\left[\frac{1}{t} + \frac{1}{c}\right] \tag{4}$$

$$= \sigma^2 \lim_{m \to \infty} m\left[\left(\sum_{i \in \mathrm{trt}} \mathbf{1}_{n_i}' \mathbf{R}_i^{-1} \mathbf{1}_{n_i}\right)^{-1} + \left(\sum_{i \in \mathrm{cont}} \mathbf{1}_{n_i}' \mathbf{R}_i^{-1} \mathbf{1}_{n_i}\right)^{-1}\right]$$

$$= \sigma^2 \lim_{m \to \infty} \left[\frac{m}{\pi \sum_{i=1}^{m} \mathbf{1}_{n_i}' \mathbf{R}_i^{-1} \mathbf{1}_{n_i}} + \frac{m}{(1 - \pi)\sum_{i=1}^{m} \mathbf{1}_{n_i}' \mathbf{R}_i^{-1} \mathbf{1}_{n_i}}\right]$$

$$= \frac{\sigma^2 m}{\pi(1 - \pi)\left(\sum_{i=1}^{m} \mathbf{1}_{n_i}' \mathbf{R}_i^{-1} \mathbf{1}_{n_i}\right)}.$$

If we assume equal cluster size $n_i = n$ for all i, then

$$V_\beta = \frac{\sigma^2}{\pi(1-\pi)\mathbf{1}_n' \mathbf{R}_i^{-1} \mathbf{1}_n}. \quad (5)$$

For the exchangeable working correlation structure, $\mathbf{1}_{n_i}' \mathbf{R}_i^{-1} \mathbf{1}_{n_i} = \frac{n_i}{1+(n_i-1)\rho}$. Thus,

$$V_\beta = \frac{\sigma^2 m}{\pi(1-\pi)\left(\sum_{i=1}^{m} \frac{n_i}{1+(n_i-1)\rho}\right)}$$

and

$$V_\beta = \frac{\sigma^2[1+(n-1)\rho]}{\pi(1-\pi)n}, \quad (6)$$

for unequal and equal cluster sizes, respectively. Please note that Eq. (6) is same as (6) in Shih's paper (Shih, 1997).

2.2. Binary outcome

For a binary outcome, we consider the logit model: $\boldsymbol{\mu}_i = \boldsymbol{p}_i = \exp(\mathbf{C}_i' \varnothing)/(1+\exp(\mathbf{C}_i' \varnothing))$. Here, $p_{ij} = p_0 = \exp(\beta_0)/(1+\exp(\beta_0))$, for all the subjects in the control group and j = 1, ⋯, n_i; and $p_{ij} = p_1 = \exp(\beta_0+\beta_1)/(1+\exp(\beta_0+\beta_1))$, for all the subjects in the treated group and j = 1, ⋯, n_i. We can show $\beta_1 = \log(\frac{p_1(1-p_0)}{(1-p_1)p_0})$, which is the logarithm of odds ratio for the response. Therefore, the null hypothesis H_0:$\beta_1 = 0$ is equivalent to $p_0 = p_1$ and the alternative hypothesis is $p_0 \neq p_1$.

Under this scenario, $\boldsymbol{D}_i = \partial \boldsymbol{\mu}_i / \partial \varnothing' = p_0(1-p_0)\left(\mathbf{1}_{n_i}, \ \mathbf{0}_{n_i}\right)$ for the subjects in the control group, denoted by $\boldsymbol{D}_{ic}$, and $\boldsymbol{D}_i = p_1(1-p_1)\left(\mathbf{1}_{n_i}, \ \mathbf{1}_{n_i}\right)$ for the subjects in the treated group, denoted by $\boldsymbol{D}_{it}$, and $\boldsymbol{V}_i = p_i(1-p_i)\boldsymbol{R}_i(\rho)$. Thus,

$$\begin{aligned}\Sigma_1 &= \sum_{i=1}^{m} \frac{1}{p_i(1-p_i)} \boldsymbol{D}_i' \mathbf{R}_i^{-1} \boldsymbol{D}_i = \sum_{i \in \text{trt}} \frac{1}{p_1(1-p_1)} \boldsymbol{D}_{it}' \mathbf{R}_i^{-1} \boldsymbol{D}_{it} + \sum_{i \in \text{cont}} \frac{1}{p_0(1-p_0)} \boldsymbol{D}_{ic}' \mathbf{R}_i^{-1} \boldsymbol{D}_{ic} \\ &= p_1(1-p_1)\left(\sum_{i \in \text{trt}} \mathbf{1}_{n_i}' \mathbf{R}_i^{-1} \mathbf{1}_{n_i}\right)\begin{pmatrix}1 & 1\\ 1 & 1\end{pmatrix} + p_0(1-p_0)\left(\sum_{i \in \text{cont}} \mathbf{1}_{n_i}' \mathbf{R}_i^{-1} \mathbf{1}_{n_i}\right)\begin{pmatrix}1 & 0\\ 0 & 0\end{pmatrix} \\ &= \begin{pmatrix}t+c & t\\ t & t\end{pmatrix},\end{aligned}$$

where $c = p_0(1-p_0)(\sum_{i \in \text{cont}} \mathbf{1}_{n_i}' \mathbf{R}_i^{-1} \mathbf{1}_{n_i})$ and $t = p_1(1-p_1)\left(\sum_{i \in \text{trt}} \mathbf{1}_{n_i}' \mathbf{R}_i^{-1} \mathbf{1}_{n_i}\right)$. Then

$$\boldsymbol{V_R} = \lim_{m \to \infty} m \begin{pmatrix} \frac{1}{c} & -\frac{1}{c} \\ -\frac{1}{c} & \frac{1}{t} + \frac{1}{c} \end{pmatrix} \quad (7)$$

And

$$V_\beta = \lim_{m \to \infty} \left[\frac{m}{p_1(1-p_1)\left(\sum_{i \in \text{trt}} \mathbf{1}'_{n_i} \mathbf{R}_i^{-1} \mathbf{1}_{n_i}\right)} + \frac{m}{p_0(1-p_0)\left(\sum_{i \in \text{cont}} \mathbf{1}'_{n_i} \mathbf{R}_i^{-1} \mathbf{1}_{n_i}\right)} \right] \quad (8)$$

$$= \lim_{m \to \infty} \left[\frac{m}{p_1(1-p_1)\pi\left(\sum_{i=1}^{m} \mathbf{1}'_{n_i} \mathbf{R}_i^{-1} \mathbf{1}_{n_i}\right)} + \frac{m}{p_0(1-p_0)(1-\pi)\left(\sum_{i=1}^{m} \mathbf{1}'_{n_i} \mathbf{R}_i^{-1} \mathbf{1}_{n_i}\right)} \right]$$

$$= \frac{m}{\sum_{i=1}^{m} \mathbf{1}'_{n_i} \mathbf{R}_i^{-1} \mathbf{1}_{n_i}} \left(\frac{1}{\pi p_1(1-p_1)} + \frac{1}{(1-\pi)p_0(1-p_0)} \right).$$

Assuming equal cluster size $n_i = n$ for all i, we have

$$V_\beta = \frac{1}{\mathbf{1}'_n \mathbf{R}_i^{-1} \mathbf{1}_n} \left(\frac{1}{\pi p_1(1-p_1)} + \frac{1}{(1-\pi)p_0(1-p_0)} \right). \quad (9)$$

When the working correlation structure is exchangeable, Eq. (8) is equivalent to the one substituting Eq. (7) into (5) in Pan's paper (Pan, 2001) and Eq. (8) derived by Shih (Shih, 1997). Additionally, Eq. (9) becomes $\frac{1+(n-1)\rho}{n}\left(\frac{1}{\pi p_1(1-p_1)} + \frac{1}{(1-\pi)p_0(1-p_0)} \right)$.

2.3. Count data

The log model is used to analyze count data, $\boldsymbol{\mu}_i = \exp(\mathbf{C}_i' \varnothing)$. $\boldsymbol{D}_i = \partial \boldsymbol{\mu}_i / \partial \varnothing' = \mathbf{C}_i \exp(\mathbf{C}_i' \varnothing)$ and $\boldsymbol{V}_i = \boldsymbol{A}_i^{1/2} \boldsymbol{R}_i(\rho) \boldsymbol{A}_i^{1/2}$, where $\boldsymbol{A_i} = \text{diag}\left\{ \exp(\mathbf{C}_{i1}' \varnothing), \cdots, \exp\left(\mathbf{C}_{in_i}' \varnothing\right) \right\}$. Specifically, $\boldsymbol{D}_{ic} = \mathbf{C}_{ic} \exp(\mathbf{C}_{ic}' \varnothing) = e^{\beta_0} \left(\mathbf{1}_{n_i}, \quad \mathbf{0}_{n_i} \right)$ for the subjects in the control group, and $\boldsymbol{D}_{it} = e^{\beta_0 + \beta_1} \left(\mathbf{1}_{n_i}, \quad \mathbf{1}_{n_i} \right)$ for the subjects in the treated group. Thus,

$$\Sigma_1 = \sum_{i \in \text{trt}} \boldsymbol{D}_{it}' \mathbf{V}_{it}^{-1} \boldsymbol{D}_{it} + \sum_{i \in \text{cont}} \boldsymbol{D}_{it}' \mathbf{V}_{it}^{-1} \boldsymbol{D}_{it}$$
$$= e^{\beta_0 + \beta_1} \left(\sum_{i \in \text{trt}} \mathbf{1}_{n_i}' \mathbf{R}_i^{-1} \mathbf{1}_{n_i} \right) \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix} + e^{\beta_0} \left(\sum_{i \in \text{cont}} \mathbf{1}_{n_i}' \mathbf{R}_i^{-1} \mathbf{1}_{n_i} \right) \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}$$
$$= \begin{pmatrix} t + c & t \\ t & t \end{pmatrix},$$

where $c = e^{\beta_0} \left(\sum_{i \in \text{cont}} \mathbf{1}_{n_i}' \mathbf{R}_i^{-1} \mathbf{1}_{n_i} \right)$ and $t = e^{\beta_0 + \beta_1} \left(\sum_{i \in \text{trt}} \mathbf{1}_{n_i}' \mathbf{R}_i^{-1} \mathbf{1}_{n_i} \right)$. Following (7),

$$V_\beta = \lim_{m \to \infty} \left[\frac{m}{e^{\beta_0 + \beta_1} \left(\sum_{i \in \text{trt}} \mathbf{1}_{n_i}' \mathbf{R}_i^{-1} \mathbf{1}_{n_i} \right)} + \frac{m}{e^{\beta_0} \left(\sum_{i \in \text{cont}} \mathbf{1}_{n_i}' \mathbf{R}_i^{-1} \mathbf{1}_{n_i} \right)} \right] \quad (10)$$
$$= \lim_{m \to \infty} \left[\frac{m}{e^{\beta_0 + \beta_1} \pi \left(\sum_{i=1}^{m} \mathbf{1}_{n_i}' \mathbf{R}_i^{-1} \mathbf{1}_{n_i} \right)} + \frac{m}{e^{\beta_0} (1 - \pi) \left(\sum_{i=1}^{m} \mathbf{1}_{n_i}' \mathbf{R}_i^{-1} \mathbf{1}_{n_i} \right)} \right]$$
$$= \frac{m}{e^{\beta_0} \left(\sum_{i=1}^{m} \mathbf{1}_{n_i}' \mathbf{R}_i^{-1} \mathbf{1}_{n_i} \right)} \left(\frac{1}{\pi e^{\beta_1}} + \frac{1}{1 - \pi} \right).$$

If we assume equal cluster size $n_i = n$ for all i, then

$$V_\beta = \frac{1}{e^{\beta_0} \mathbf{1}_n' \mathbf{R}_i^{-1} \mathbf{1}_n} \left(\frac{1}{\pi e^{\beta_1}} + \frac{1}{1 - \pi} \right). \quad (11)$$

With the assumptions of exchangeable working correlation structure, equal cluster size n and equal allocation $\pi = 0.5$, we have $V_\beta = \frac{2[1 + (n-1)\rho]}{n e^{\beta_0}} \left(\frac{1}{e^{\beta_1}} + 1 \right)$. It is Eq. (19) developed by Bhaumik and Gibbons (Amatya, Bhaumik, et al., 2013).

3. Relative efficiency of unequal versus equal cluster sizes

We will use relative variance of the treatment effect to define RE when comparing unequal to equal cluster sizes. Let Ω_{equal} denote a design with equal cluster sizes n, and let Ω_{unequal} denote a design with unequal cluster sizes n_i, $i = 1, \cdots, m$. Obviously $mn = \sum_{i=1}^{m} n_i$. The RE of unequal versus equal cluster sizes for the treatment effect, $\text{RE}(\widehat{\beta_1})$, is defined as

$$\mathrm{RE}\left(\widehat{\beta_1}\right) = \frac{\mathrm{Var}(\widehat{\beta_1}|\Omega_{\mathrm{equal}})}{\mathrm{Var}(\widehat{\beta_1}|\Omega_{\mathrm{unequal}})}. \quad (12)$$

From equations (4) and (5) for a continuous outcome, (8) and (9) for a binary outcome, and (10) and (11) for count data, we have the same

$$\mathrm{RE}\left(\widehat{\beta_1}\right) = \frac{\sum_{i=1}^{m} \mathbf{1}'_{n_i} \mathbf{R}_i^{-1} \mathbf{1}_{n_i}}{m \mathbf{1}'_n \mathbf{R}_i^{-1} \mathbf{1}_n}. \quad (13)$$

REs are surprisingly kept the same for all three types of outcomes. In other words, REs are independent of the outcome type provided that GEE models are used.

With the assumption of exchangeable working correlation structure,

$$\mathbf{1}'_{n_i} \mathbf{R}_i^{-1} \mathbf{1}_{n_i} = \frac{n_i}{1 + (n_i - 1)\rho},$$

then Eq. (13) equals

$$\mathrm{RE}\left(\widehat{\beta_1}\right) = \frac{1 + (n-1)\rho}{n} \frac{1}{m} \sum_{i=1}^{m} \frac{n_i}{1 + (n_i - 1)\rho}. \quad (14)$$

Let $\tau = (1 - \rho)/\rho$, Eq. (14) becomes

$$\mathrm{RE}\left(\widehat{\beta_1}\right) = \frac{n + \tau}{n} \frac{1}{m} \sum_{i=1}^{m} \frac{n_i}{n_i + \tau}. \quad (15)$$

It is identical to Eq. (9) developed by Breukelen et al.(Van Breukelen, Candel, et al., 2007). For binary outcomes, the same finding is shown in equation (A8) (Candel and Van Breukelen, 2010) and section 7 (Van Breukelen et al., 2015). They show that

1. There is a trade-off between mean cluster size n and τ. Multiplying all n_i's and τ by a factor >0 does not change the $\mathrm{RE}\left(\widehat{\beta_1}\right)$.
2. As $\rho \to 0$ or as $\rho \to 1$, the $\mathrm{RE}\left(\widehat{\beta_1}\right) \to 1$, for any cluster size distribution. For $0<\rho<1$, the $\mathrm{RE}\left(\widehat{\beta_1}\right)$ is smaller than 1 and so the equal cluster size design is optimal.

4. Simulation studies

In order to compute RE$(\widehat{\beta_1})$ in Eq. (14), Breukelen et al (Van Breukelen, Candel, et al., 2007) proposed a uniform, positively skewed, negatively skewed, bimodal and unimodal distribution of cluster size, respectively. Let X_i be random variable of cluster size n_i. Here, n_i, $i = 1, \ldots, m$ are dependent since they must sum to $N = mn$. We assume that cluster sizes $(X_1, \cdots, X_m)$ follow a multinomial distribution with N and probabilities $(p_1, \cdots, p_m)$, where $\sum_{i=1}^{m} p_i = 1$. Under this distribution, the mean of X_i is Np_i, the variance is $Np_i(1 - p_i)$ and covariance of (X_i, X_j) is $-Np_ip_j$ for $i \neq j$. Given the number of clusters m and equal cluster size n, the total sample size is $N = mn = \sum_{i=1}^{m} n_i$ under the designs of Ω_{equal} and Ω_{unequal}.

For convenience of the following discussion, we sort the i-th cluster with the distribution probability p_i's, $i = 1, \cdots, m$, by a non-decreasing order such that $p_1 \leq p_2 \leq \cdots \leq p_m$. Obviously, the cluster sizes n_i's, $i = 1, \cdots, m$ may be different even if p_i's are equal. Therefore, six patterns of $(p_1, \cdots, p_m)$ are discussed under the design Ω_{unequal}.

1. Constant: $p_1 = p_2 \cdots = p_m$
2. Monotonically increasing: $p_1 < p_2 < \cdots < p_m$;
3. Constant followed by monotonically increasing: $p_1 = p_2 \cdots = p_k < \cdots < p_m$;
4. Monotonically increasing followed by constant: $p_1 < p_2 \cdots < p_k = \cdots = p_m$;
5. Constant, monotonically increasing followed by constant: $p_1 = p_2 \cdots = p_{k_1} < p_{k_1+1} < \cdots < p_{k_2} = \cdots = p_m$;
6. Monotonically increasing, constant followed by monotonically increasing: $p_1 < p_2 \cdots < p_{k_1} = p_{k_1+1} = \cdots = p_{k_2} < \cdots < p_m$;

These six patterns are shown in Figure 1, which demonstrates the probabilities p_i, $i = 1, \cdots, 100$. We sort these 100 probabilities by a non-decreasing order, such that the 1st cluster has the minimum of probabilities and the 100th cluster has the maximum of the probabilities. For example, pattern 2, the probabilities of 1st cluster and 100th cluster are 0.006 and 0.013, respectively; they are 0.004 and 0.0275 for pattern 3. There could be more complicated patterns but they should be combinations of the above six patterns. Therefore, RE$(\widehat{\beta_1})$ in Eq. (14) will be computed through simulation studies only for these six patterns.

4.1. Designs

As shown in Eq. (14), RE$(\widehat{\beta_1})$ is independent of cluster allocations π and the parameters β_0 and β_1. To investigate the RE based on GEE models, the following factors are considered: (1) number of clusters m and equal cluster sizes n (or N); (2) the values of $p_1, \cdots, p_m$, equivalently, the pattern of $p_1, \cdots, p_m$; (3) association parameter (ρ) in the exchangeable working correlation structure.

Number of clusters m and equal cluster sizes n results in total sample size of $N = nm$. We will consider three scenarios: small, medium, and large CRTs; eg. $N = 200, 600, 2000,$

respectively. Within each scenario we investigate three situations: small, medium, and large number of clusters. For example, $m = 20$ and $n = 100$; $m = 50$ and $n = 40$; $m = 100$ and $n = 20$ in a large CRT with a total sample size of $N = 2000$. All six patterns in Figure 1 are used for each study design.

Table 1 shows the values of $(p_1, \cdots, p_m)$ for the simulation design with $m = 100$ and $n = 20$. During the interval of monotonically increasing within each pattern, we make $p_i's$ as an arithmetic sequence for code efficiency. All the values of p_i, $i = 1, \cdots, m$ are shown in the 3rd column for each pattern. For example, pattern 2, $p_i = p_1 + (i - 1)d$, $i = 2, \cdots, m$. Once p_1 is given, d will be calculated from p_1 and m and the formula is shown in the last column. Consequentially, p_i, $i = 2, \cdots, m$ will be computed from the formula in the 3rd column. Pattern 3 has the probabilities of p_i with constant followed by monotonically increasing: $p_1 = \cdots = p_k < p_{k+1} < \cdots < p_m$. In this pattern, k and P_k are pre-determined. Here $k = 50$ and $P_k = 0.2$ are used, respectively. Then p_i, $i = 1, \cdots, k$ and d will be computed from the formulas in the 3rd column and the last column, respectively. Finally, p_i, $i = k + 1, \cdots, m$ will be calculated from the formula in the 3rd column. The logic for pattern 4 is the same as the one for pattern 2. For pattern 5, the pre-determined values are k_1, k_2, and P_{k_1}. Then p_i, $i = 1, \cdots, k_1$ and d will be computed from the formula in the 3rd column and the last column, respectively. Finally, p_i, $i = k_1 + 1, \cdots, k_2$ will be calculated from the formula in the 3rd column. Please note $p_{k_2} = \cdots = p_m$, thus, all the values of p_i, $i = 1, \cdots, m$ are obtained. Pattern 6 is the most complicated case. Four parameters including k_1, k_2, P_{k_1}, and P_{k_2} must be pre-determined. Then p_i, $i = k_1, \cdots, k_2$ and d_i, $i = 1, 2$ will be computed from the formula in the 3rd column and the last column, respectively. Finally, p_i, $i = 1, \cdots, k_1 - 1, k_2 + 1, \cdots, m$ will be calculated from the formula in the 3rd column. The required parameters are marked with # in table 1. Provided the values of $(p_1, \cdots, p_m)$ are available for each pattern, we will simulate the cluster sizes n_i, $i = 1, \cdots, m$ from a multinomial distribution with N and probabilities $(p_1, \cdots, p_m)$.

For all the scenarios, the required parameters for $(p_1, \cdots, p_m)$ calculation are shown in Table 2. Even if the ICCs are pretty small for most CRTs (Murray, Varnell, et al., 2004; Turner, Prevost, et al., 2004), the association parameter (ρ) ranged from 0 to 0.95 with steps of 0.01 considered for the illustration purposes. 1000 simulation samples are generated for each design. REs are calculated using Eq. (14) for all the samples, and mean, standard deviation, minimum and maximum of REs are obtained at each ρ correspondingly. Source code to reproduce the results is available as Supporting Information on the journal's web page (http://onlinelibrary.wiley.com/doi/xxx/suppinfo).

4.2. Results

Figure 2 shows the plots of the mean, minimum and maximum of RE as a function of association parameter (ρ) based on the simulation data with $m = 100$ and $n = 20$ for patterns 1–6, respectively. The parameters used in this simulation study are shown in Table 1. As can be seen, the mean RE starts from 1, reaches the minimum, and then increases to 1 at ρ=1. The minimum of mean REs for these six patterns are 0.9871 at ρ=0.05, 0.9734 at ρ=0.05, 0.8589 at ρ=0.05, 0.9343 at ρ=0.08, 0.8962 at ρ=0.07, and 0.9570 at ρ=0.06, respectively.

The standard deviations are so small (<0.008) for all the ρ's and all six patterns, thus we do not include 95% CIs in the plots. When comparing these six patterns, we find that the first pattern shows the highest RE. This is expected since the setting of $p_1 = p_2 = \cdots = p_m$ has the highest chance of equal cluster sizes. The second pattern has the next highest RE while the third pattern shows the lowest one. Pattern 6 has a better RE than patterns 4 and 5. Simultaneously, after the minimum is reached, the first pattern is the least steep, while the third one is very steep, and pattern 5 is a little steeper than pattern 6. The closer to ρ that the minimum of mean REs is reached, the larger range RE has. The same findings are observed across all six patterns. The plots of the mean and range of RE based on the simulation data with $m = 20$, $n = 100$ and $m = 50$, $n = 40$ for six patterns are shown in Appendix Figure 1 and 2, respectively. Among three situations including small, medium, and large number of clusters, we notice that the range looks narrower when the cluster size increases at any value of ρ for any pattern. For a small cluster size ($n = 20$), the range of mean RE stays large when ρ increases after the minimum of mean REs is reached for patterns 2–6. There are not many differences otherwise.

Table 3 presents mean (min, max) of CV, the minimum and median of mean RE and corresponding ρ's among 1000 repeated samples for all three total sample sizes (200, 600, and 2000), the three situations of cluster sizes and number of clusters, and all six patterns. The first pattern indeed shows the highest RE for any scenario. Below discussions focus on the remaining five patterns.

1. Total sample size is small (N=200); the minimums of mean REs are less than 90% for one-third of cases. When the number of clusters is small ($m = 5$), the minimums of mean RE for pattern 4 and pattern 5 are 0.9868 and 0.7974, respectively, the largest and smallest estimates among the five patterns. Additionally, the minimums are reached at $\rho \in [0.02, 0.06]$. The remaining two situations ($m = 20, 40$). obtain the largest estimate of minimum of mean RE for pattern 3 and the smallest estimate for pattern 5. We also find that the minimum is reached at the larger ρ when cluster size decreases. For example, ρ is around 0.05 for $n = 40$, around 0.15 for $n = 20$, and around 0.25 for $n = 5$, respectively. The median of mean REs are at least 93% for $m = 5$ and $n = 40$, 93% for $m = 20$ and $n = 10$, 88% for $m = 40$ and $n = 5$. The corresponding values of ρ are between 0.48 and 0.49, between 0.49 and 0.53, and between 0.53 and 0.60, respectively.

2. Total sample size is medium (N=600); All the minimums of mean REs are at least 84% and more than half are above 90%. When the number of clusters is small ($m = 6$), the minimums of mean RE for pattern 5 and pattern 6 are 0.8450 and 0.9315, respectively. They are the largest and smallest estimates among patterns 2–6. The remaining two situations ($m = 30, 60$) obtain the smallest estimate of minimum of mean RE for pattern 3. We have the same findings about the trend of ρ, where the minimum is reached, for different cluster sizes: the smaller the cluster size, the larger the ρ. Additionally, all the medians of mean REs are at least 95% for $m = 6$ and $n = 100$, 96% for $m = 30$ and $n = 20$, and 92% for $m = 60$ and $n = 10$, respectively. The corresponding values of ρ are

between 0.48 and 0.49, between 0.48 and 0.50, and between 0.49 and 0.52, respectively.

3. Total sample size is large (N=2000); Such large studies will definitely yield sufficient power. They have the best estimate of minimal REs with at least 85% for all the different settings (m, n, and pattern) and approximately 75% cases are above 90%. The situations (m = 20, 50) obtain the largest estimate of minimum of RE for pattern 6. Pattern 3 demonstrates the smallest estimate of minimum mean RE for all three situations (87.38%, 86.67%, 85.89%). ρ, where the minimum is reached, ranged from 0.01 to 0.08. All the medians of mean REs are larger than 99% for m = 20 and n = 100, above 98% for m = 50 and n = 40, and at least 96% m = 100 and n = 20.

We find that minimum of mean RE decreases when CV increases for all simulation studies. Pattern 1 indicates the lowest CV as expected while pattern 5 has the largest CV in most cases. In summary, the worst scenarios for small CRTs give 21%, 11%, and 14% efficiency loss when the number of cluster is small, medium and large, respectively. They are 16%, 14%, and 16% for medium CRTs. For large CRTs all three are about the same, 14%.

Additionally, Table 3 shows that ρ at which the RE is minimum decreases with increasing average cluster size n given a pattern. Breukelen et al (Van Breukelen, Candel, et al., 2007) shows the RE minimum is reached at $\rho = 1/(n + 1)$ based on their Taylor approximation (Eq. (10)). We have the similar results for large CRTs with mn = 2000, especially for n = 40 and 60. Furthermore, Breukelen et al (Van Breukelen, Candel, et al., 2007) shows that the number of clusters does not affect RE, Eq. (15), given the cluster distribution. We make m twice and keep n same in the simulation designs. The parameters for (p_1, ⋯, p_m) and simulation results are shown in Appendix Tables 1 and 2. When we compare Table 3 and Appendix Table 2, e.g. m = 5, n = 40 and m = 10, n = 40, the different findings within a same pattern are noticed. For example, pattern 2, the minimums of mean REs are 0.9617 and 0.8802, respectively. On the other hand, they have different CVs, 0.42 (0.17, 0.69) and 0.65 (0.52, 0.78). That is, the same patterns in our simulation setting which strongly depends on (p_1, ⋯, p_m) may have the different cluster distributions when the number of cluster doubles.

5 Sample size calculations

In sections 5.1 and 5.2, we show the sample size formulas for the different types of outcome based on GEE models with the assumption of equal cluster sizes with no cost constraints and under a given cost, respectively. Section 5.3 demonstrates how the sample sizes are adjusted due to efficiency loss from unequal cluster sizes.

5.1. No cost constraints

For continuous and binary count outcomes, a two-sided test based on β with type I error of α and nominal power η requires the sample size

$$m = \left(z_{1-\alpha/2} + z_{\eta}\right)^2 \frac{V_{\beta}}{\beta^2}, \quad (16)$$

Where β is the treatment effect. Substituting Equations (6) and (9) for a two-group study with the assumption of exchangeable correlation matrix, we have the sample size formulas

$$m = \left(z_{1-\alpha/2} + z_{\eta}\right)^2 \frac{\sigma^2[1 + (n-1)\rho]}{\pi(1-\pi)n\beta^2}, \quad (17)$$

and

$$m = \left(z_{1-\alpha/2} + z_{\eta}\right)^2 \frac{\left(\frac{1}{\pi p_1(1-p_1)} + \frac{1}{(1-\pi)p_0(1-p_0)}\right)[1 + (n-1)\rho]}{n\left[\log\left(\frac{p_1(1-p_0)}{p_0(1-p_1)}\right)\right]^2},$$

respectively, where z_a is the $100 \times a$ percentile of a standard normal distribution. Please note Eq. (17) reduces to Eq. (3) (Van Breukelen and Candel, 2012) when the equal allocation occurs.

For count outcomes, we use formula (3) (Amatya, Bhaumik, et al., 2013) to obtain

$$m = \left(z_{1-\alpha/2}\sqrt{\frac{1}{\pi} + \frac{1}{1-\pi}} + z_{\eta}\sqrt{\frac{1}{\pi e^{\beta}} + \frac{1}{1-\pi}}\right)^2 \frac{1 + (n-1)\rho}{n e^{\beta_0}\beta^2}$$

through Eq. (11).

5.2. Optimal design

Van Breukelen et al (Van Breukelen and Candel, 2015) define the total budget B in CRTs. Assume the study cost per cluster in the enrollment stage is c currency units (e.g. USD), and each subject costs s currency units. Total budget B includes m clusters and n subjects within a cluster,

$$B = m(c + sn). \quad (18)$$

They define “optimal” as that the variance of the treatment effect estimator in the linear regression model is minimized under the cost constraints. In this paper the term “optimal” using GEE models refers to maximum power for a given sampling budget. Actually, they are equivalent given an unbiased point estimator of the treatment effect β. Here, the optimal design is to find the design which maximizes the power given the constraint in Eq. (18).

Define

$$\Delta_1 = \frac{\sigma^2}{\pi(1-\pi)\beta^2}$$

for continuous outcome, and

$$\Delta_1 = \frac{\frac{1}{\pi p_1(1-p_1)} + \frac{1}{(1-\pi)p_0(1-p_0)}}{\left[\log\left(\frac{p_1(1-p_0)}{p_0(1-p_1)}\right)\right]^2}$$

for binary outcome, the sample sizes based on GEE models (Shih, 1997; Pan, 2001) can be re-written as

$$m = \left(z_{1-\alpha/2} + z_\eta\right)^2 \frac{1+(n-1)\rho}{n} \quad \Delta_1 . \quad (19)$$

Alternatively,

$$z_\eta = \sqrt{\frac{n}{1+(n-1)\rho}\frac{m}{\Delta_1}} - z_{1-\alpha/2} .$$

For count data,

$$z_\eta = \left(\sqrt{\frac{n}{1+(n-1)\rho} m e^{\beta_0} \beta^2} - z_{1-\alpha/2}\sqrt{\frac{1}{\pi}+\frac{1}{1-\pi}}\right) \Big/ \sqrt{\frac{1}{\pi e^{\beta}} + \frac{1}{1-\pi}} . \quad (20)$$

Consequently, the power is calculated by $\phi^{-1}(z_\eta)$, where ϕ is cumulative density function of standard normal distribution.

As noted that π, α, Δ_1 in Eq. (19) and (β_0, β) in Eq. (20) are pre-determined values under null and alternative hypothesis for three types of outcomes, maximizing the power means maximizing

$$K = \frac{nm}{1+(n-1)\rho} .$$

For a known value ρ, the locally optimal design (LOD) is reached at

$$n_{\mathrm{LOD}} = \sqrt{\frac{\vartheta c}{s}}, \qquad m_{\mathrm{LOD}} = \frac{B}{\sqrt{\vartheta s c} + c}, \quad (21)$$

Where $\vartheta = \frac{1-\rho}{\rho}$. It is identical to the optimal designs proposed by other researchers (Raudenbush, 1997; Moerbeek et al., 2000; Van Breukelen and Candel, 2015), even though they use the different definition of “optimal.”

In the scenario of an unknown parameter value of ρ, Van Breukelen and Candel (Van Breukelen and Candel, 2015) proposed the Maximin designs (MMDs) using the linear regression model and the variance of the maximum likelihood estimator. However, there is no closed form for binary outcomes and the variance from binary outcomes needs to be transformed to that of continuous outcomes. When using GEE models, Liu et al (Liu et al., 2017) proposed an algorithm, shown as below, to find the optimal design (m_{OD}, n_{OD}) based on parameter space (ρ_{min}, ρ_{max}) and design space (m_{min}, m_{max}).

Step 1, Define the parameter and design space, (ρ_{min}, ρ_{max}) and (m_{min}, m_{max}), respectively.

Step 2, Consider $\rho = \rho_{max}$, calculate m_{LOD} using equation (21).

A. If it is within the range (m_{min}, m_{max}), then set $m_{OD} = m_{LOD}$.

B. If it is outside of (m_{min}, m_{max}), then set $m_{OD} = m_{max}$. The cluster size n_{OD} is calculated by $\left(\frac{B}{m_{OD}} - c\right)/s$.

Liu et al (Liu and Colditz, 2017) shows proof details of LODs and ODs when GEE models are considered in the sample size calculation. When comparing the algorithm in MMDs based on RE and efficiency (Van Breukelen and Candel, 2015), the revised algorithm is same as one of their maximin designs which maximizes the minimum efficiency. However, it works for any type of outcome including continuous, binary, and count.

Given a fixed budget, we can find the optimal design which maximizes the power for the proposed study but please note that the maximized power may be much different from nominal power η, eg. 80%. When a known value ρ is the same, or the parameter and design space are the same for three different types of outcomes, the optimal sample sizes including number of clusters and cluster size are same for these three as well. However, the maximized powers are different and depends on the outcome type.

5.3. Application

Equal cluster sizes are assumed in sections 5.1 and 5.2. However, equal cluster sizes are not guaranteed in practice. When designing a CRT to examine the treatment effect through GEE models, first we assume equal cluster sizes and n is given. The required number of clusters can be obtained from these two sections. We will then increase number of clusters in order to compensate for the efficiency loss due to unequal cluster sizes. Below are examples to demonstrate the calculation of number of clusters for unequal cluster sizes.

If we consider GEE models to design a CRT without financial constraints, the sample size formulas in section 5.1 are used for calculation. For instance, we want to compare two physical therapy treatments designed to increase muscle flexibility in a CRT. We must determine how many clusters are required to achieve a power of at least 80% to detect a group mean difference with $n = 40$ at the type I error of 5%. The mean difference of muscle flexibility between the standard treatment and the new treatment is 1, the common standard deviation of muscle flexibility is 3.1, and equal allocation ($\pi = 0.5$) and $\rho = 0.01$ are considered. Using equation (17), the number of clusters $m = 20$ are needed. The total sample

size is $mn = 800$. We consider this study as a medium CRT with medium number of clusters. As shown in section 4.2, REs are dependent of the pattern of $p_1, \cdots, p_m$ and the worst scenarios for medium CRTs give 16%, 14%, and 16% efficiency loss for medium CRTs when the number of cluster is small, medium and large, respectively. The adjusted number of clusters will be 20/0.86=24.

If a study is designed under a given budget, then the sample size formulas in section 5.2 will be used. For example, total budget of the proposed study is $55,000, the study cost per cluster in the enrollment stage is $1,000, and each subject costs $100. Using equation (21), the locally optimal design with $n_{\mathrm{LOD}} \approx 8.00$ and $m_{\mathrm{LOD}} \approx 30.55$ are obtained given ρ=0.135. It is considered a small CRT with large number of clusters. The worst scenario for small CRTs gives 14% efficiency loss when the cluster size is small. Therefore, the adjusted number of clusters will be $30.55/0.86 \approx 35.5$. We need to enroll 36 clusters with cluster size of 8. The original proposed budget is 30(1000+100*8)=$54,000 and is within the budget while the budget after the adjustment is 36*1000+100*36*8=$64,800 and is beyond the budget. Obviously, the cost could be increased because the number of clusters rise. Therefore, there is a trade-off between the increase in number of clusters due to efficiency loss and the actual cost within the budget in the optimal design.

6 Discussion

The sample size formulas based on GEE methods have been derived in recent years (Liu and Liang, 1997; Shih, 1997; Pan, 2001; Amatya, Bhaumik, et al., 2013). As we notice, the sample size formulas assume the exchangeable working correlation structure and equal cluster sizes. However, the assumption of equal cluster size is not realistic. Therefore, researchers investigate the RE of unequal versus equal cluster sizes when testing the treatment effect. In the paper (Van Breukelen, Candel, et al., 2007), Breukelen et al consider linear regression models to address the RE of unequal versus equal cluster sizes with continuous outcome for CRTs. For a binary outcome in CRTs, Candel and Breukelen (Candel and Van Breukelen, 2010) utilize mixed effect logistic regression model and present the adjusted sample size formula for varying cluster sizes. In this paper, we investigate the REs based on GEE models to test the treatment effect in a two-group comparison in CRTs. The three outcomes of continuous, binary, or count data are discussed simultaneously. The variances of the estimator of the treatment effect were derived for three different types of outcome given the exchangeable working correlation structure. We define RE as the ratio of variance of the estimator of the treatment effect for equal to unequal cluster sizes. The simpler formulas of RE with continuous, binary, and count outcomes are given.

First, we find the formulas of REs are the same for continuous, binary, and count data using GEE models. That is, RE is not dependent on the type of outcome at the stage of study design. Second, $\mathrm{RE}(\widehat{\beta_1})$ is independent of cluster allocations $\boldsymbol{\pi}$ and the parameters β_0 and β_1.

Third, the performance of RE for the exchangeable working correlation structure is investigated through simulations since no closed form is provided. We make a conclusion about efficiency loss for the different sizes of CRTs-small, medium, and large. Fourth, the adjusted number of clusters accounting for efficiency loss are illustrated for CRTs without financial constraints and under a budget. Fifth, the expressions (Eq. (14) and (15)) are same

as those (Eq. (8) and (9)) in Breukelen et al (Van Breukelen, Candel, et al., 2007). They are an extension of published results on optimal sample sizes and on the relative efficiency of unequal versus equal cluster sizes in cluster randomized trials using GEE models. Last, Breukelen et al (Van Breukelen, Candel, et al., 2007) approximate the RE through Taylor series by assuming that n_i, $i = 1, \cdots, m$ are independent realizations of a random variable while we conduct simulation studies to investigate RE performance given that unequal cluster sizes follow a multinomial distribution under $mn = \sum_{i=1}^{m} n_i$. The rationale here is that the number of clusters under the assumption of equal cluster sizes in the sample size calculation are obtained, resulting in a fixed sample size ($N = mn$). Next we investigate the performance of RE for unequal cluster size n_i. Breukelen et al (Van Breukelen, Candel, et al., 2007) find that the loss of efficiency could be larger than 10% when CV exceeds 0.63. However, such a large CV is not expected to occur in real cluster randomized trials. As such, they conclude that the loss of efficiency due to variation of cluster sizes rarely exceeds 10 per cent and can be compensated by sampling 11 per cent more clusters. Candel and Breukelen (Candel and Van Breukelen, 2010) find that 14% more clusters are needed to cover the efficiency loss in many cases with a binary outcome in CRTs. Our simulation results show that small CRTs give 21%, 11%, and 14% efficiency loss when the number of cluster is small, medium and large, respectively. They are 16%, 14%, and 16% for medium CRTs. The efficiency loss could be approximately 14%, even in large CRTs under the fact of $mn = \sum_{i=1}^{m} n_i$. Please note that the corresponding CVs are larger than 0.63, thus, we make the similar conclusion as Breukelen et al (Van Breukelen, Candel, et al., 2007).

There are several limitations to this approach. The first limitation is that the covariates are not included in the consideration. If we include the covariates in GEE model, the sample size formulas are not shown to be simple as sections 5.1 and 5.2. Moreover, the performance of the sample size formula is sensitive to the distribution of the covariates (Liu and Liang, 1997). However, it is reasonable to calculate the sample size without the consideration of covariates in the design stage in general. The next limitation is that our proposed RE is investigated on the exchangeable working correlation matrix only. The assumption of the exchangeable working correlation matrix may not hold in a real world application. However, the sample size formulas always use this assumption and it is acceptable in practice. The third limitation is that the performance of RE is determined by the pattern of $p_1, \cdots, p_m$ as well as the number of cluster, cluster sizes, and association parameter (ρ). Obviously, we have no prior knowledge of $p_1, \cdots, p_m$. This paper shows the different scenarios of efficiency loss from six basic patterns. Sample size adjustments are made based on the worst scenario in order to be conservative. The last limitation is that GEE models may have poor performance when the number of clusters is not large enough. The performance of RE defined in section 3 and the sample size formula in section 5 must be used cautiously.

In conclusion, this paper discusses the efficiency loss based on GEE models when unequal cluster sizes occur in a real world. We believe that the proposed method is very useful and practical, especially for designing CRTs with any type of outcomes.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements

The authors wish to express sincere thanks to associate editor and two reviewers for their constructive and valuable comments and suggestions which considerably improved the manuscript. We thank the Alvin J. Siteman Cancer Center at Washington University School of Medicine and Barnes-Jewish Hospital in St. Louis, MO., for supporting this research (P30 CA91842).

References

Amatya A, Bhaumik D and Gibbons R (2013). Sample size determination for clustered count data. Statistics in Medicine 32:4162–4179. [PubMed: 23589228]

Austin P (2007). A comparison of the statistical power of different methods for the analysis of cluster randomization trials with binary outcomes. Statistics in Medicine 26:3550–3565. [PubMed: 17238238]

Campbell MK, Mollison J, Steen N, Grimshaw JM and Eccles M (2000). Analysis of cluster randomized trials in primary care: a practical approach. Family practice 17(2):192–196. [PubMed: 10758085]

Candel MJ and Van Breukelen GJ (2010). Sample size adjustments for varying cluster sizes in cluster randomized trials with binary outcomes analyzed with second-order PQL mixed logistic regression. Stat Med 29(14):1488–1501. [PubMed: 20101669]

Candel MJJM, Van Breukelen GJP, Kotova L and Berger MPF (2008). Optimality of Equal vs. Unequal Cluster Sizes in Multilevel Intervention Studies: A Monte Carlo Study for Small Sample Sizes. Communications in Statistics - Simulation and Computation 37(1):222–239.

Donner A, Birkett N and Buck C (1981). Randomization by cluster-sample size requirements and analysis. Am J Epidemiol 114:906–914. [PubMed: 7315838]

Eldridge S, Ashby D and Kerry S (2006). Sample size for cluster randomized trials: effect of coefficient of variation of cluster size and analysis method. Int J Epidemiol 35:1292–1300. [PubMed: 16943232]

Gravenstein S, Dahal R, Gozalo PL, Davidson HE, Han LF, Taljaard M and Mor V (2016). A cluster randomized controlled trial comparing relative effectiveness of two licensed influenza vaccines in US nursing homes: Design and rationale. Clinical trials (London, England).

Gulliford MC, van Staa TP, McDermott L, McCann G, Charlton J and Dregan A (2014). Cluster randomized trials utilizing primary care electronic health records: methodological issues in design, conduct, and analysis (eCRT Study). Trials 15:220. [PubMed: 24919485]

Jeffe DB, Perez M, Cole EF, Liu Y and Schootman M (2016). The Effects of Surgery Type and Chemotherapy on Early-Stage Breast Cancer Patients' Quality of Life Over 2-Year Follow-up. Annals of surgical oncology 23(3):735–743. [PubMed: 26511265]

Kalfon P, Mimoz O, Loundou A, Geantot MA, Revel N, Villard I, Amour J, Azoulay E, Garrouste-Orgeas M, Martin C, Sharshar T, Baumstarck K and Auquier P (2016). Reduction of self-perceived discomforts in critically ill patients in French intensive care units: study protocol for a cluster-randomized controlled trial. Trials 17(1):87. [PubMed: 26880373]

Liang K-Y and Zeger SL (1986). Longitudinal Data Analysis Using Generalized Linear Models. Biometrika 73(1):13–22.

Lin CC, Bruinooge SS, Kirkwood MK, Olsen C, Jemal A, Bajorin D, Giordano SH, Goldstein M, Guadagnolo BA, Kosty M, Hopkins S, Yu JB, Arnone A, Hanley A, Stevens S and Hershman DL (2015). Association Between Geographic Access to Cancer Care, Insurance, and Receipt of Chemotherapy: Geographic Distribution of Oncologists and Travel Distance. Journal of clinical oncology : official journal of the American Society of Clinical Oncology 33(28):3177–3185. [PubMed: 26304878]

Liu G and Liang K-Y (1997). Sample Size Calculations for Studies with Correlated Observations. Biometrics 53(3):937–947. [PubMed: 9290224]

Liu J and Colditz GA (2017). Optimal design of longitudinal data analysis using generalized estimating equation models. Biometrical journal. Biometrische Zeitschrift 59(2):315–330. [PubMed: 27878852]

Manatunga AK, Hudgens MG and Chen S (2001). Sample Size Estimation in Cluster Randomized Studies with Varying Cluster Size. Biometrical Journal 43(1):75–86.

Mehring M, Haag M, Linde K, Wagenpfeil S and Schneider A (2016). Effects of a Web-Based Intervention for Stress Reduction in Primary Care: A Cluster Randomized Controlled Trial. Journal of medical Internet research 18(2):e27. [PubMed: 26872703]

Moerbeek M, Van Breukelen G and Berger M (2000). Design Issues for Experiments in Multilevel Populations. Journal of Educational and Behavioral Statistics 25(3):271–284.

Murray DM, Varnell SP and Blitstein JL (2004). Design and analysis of group-randomized trials: a review of recent methodological developments. American journal of public health 94(3):423–432. [PubMed: 14998806]

Nagayama H, Tomori K, Ohno K, Takahashi K, Ogahara K, Sawada T, Uezu S, Nagatani R and Yamauchi K (2016). 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 11(3):e0150374. [PubMed: 26930191]

Pan W (2001). Sample size and power calculations with correlated binary data. Control Clin Trials 22(3):211–227. [PubMed: 11384786]

Park YH, Jung KH, Im SA, Sohn JH, Ro J, Ahn JH, Kim SB, Nam BH, Oh do Y, Han SW, Lee S, Park IH, Lee KS, Kim JH, Kang SY, Lee MH, Park HS, Woo SY, Jung SH, Ahn JS and Im YH (2015). 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 research and treatment 152(1):77–85. [PubMed: 26033708]

Raudenbush S (1997). Statistical analysis and optimal design for cluster randomized trials. Psychol Methods 2:173–185.

Rosner B and Glynn R (2011). Power and Sample size estimation for the clustered Wilcoxon test. Biometrics 67:646–653. [PubMed: 20880011]

Rosner B, Glynn R and Lee M (2003). Incorporation of clustering effects for the Wilcoxon rank sum test: a large-sample approach. Biometrics 59:1089–1098. [PubMed: 14969489]

Sanda MG, Dunn RL, Michalski J, Sandler HM, Northouse L, Hembroff L, Lin X, Greenfield TK, Litwin MS, Saigal CS, Mahadevan A, Klein E, Kibel A, Pisters LL, Kuban D, Kaplan I, Wood D, Ciezki J, Shah N and Wei JT (2008). Quality of life and satisfaction with outcome among prostate-cancer survivors. The New England journal of medicine 358(12):1250–1261. [PubMed: 18354103]

Shih WJ (1997). Sample Size and Power Calculations for Periodontal and Other Studies with Clustered Samples Using the Method of Generalized Estimating Equations. Biometrical Journal 39(8):899–908.

Teerenstra S, Lu B, Preisser JS, van Achterberg T and Borm GF (2010). Sample size considerations for GEE analyses of three-level cluster randomized trials. Biometrics 66(4):1230–1237. [PubMed: 20070297]

Toriola AT, Liu J, Ganz PA, Colditz GA, Yang L, Izadi S, Naughton MJ, Schwartz AL and Wolin KY (2015). Effect of weight loss on bone health in overweight/obese postmenopausal breast cancer survivors. Breast cancer research and treatment 152(3):637–643. [PubMed: 26175059]

Turner RM, Prevost AT and Thompson SG (2004). Allowing for imprecision of the intracluster correlation coefficient in the design of cluster randomized trials. Statistics in medicine 23(8):1195–1214. [PubMed: 15083478]

Van Breukelen G and Candel M (2015). Efficient design of cluster randomized and multicentre trials with unknown intraclass correlation. Stat Methods Med Res 24(5):540–556. [PubMed: 21937473]

Van Breukelen GJ and Candel MJ (2012). Calculating sample sizes for cluster randomized trials: we can keep it simple and efficient! Journal of clinical epidemiology 65(11):1212–1218. [PubMed: 23017638]

Van Breukelen GJ, Candel MJ and Berger MP (2008). Relative efficiency of unequal cluster sizes for variance component estimation in cluster randomized and multicentre trials. Stat Methods Med Res 17(4):439–458. [PubMed: 17698940]

Van Breukelen GJP, Candel MJJM and Berger MPF (2007). Relative efficiency of unequal versus equal cluster sizes in cluster randomized and multicentre trials. Statistics in Medicine 26(13): 2589–2603. [PubMed: 17094074]

Yamagata K, Makino H, Iseki K, Ito S, Kimura K, Kusano E, Shibata T, Tomita K, Narita I, Nishino T, Fujigaki Y, Mitarai T, Watanabe T, Wada T, Nakamura T and Matsuo S (2016). Effect of Behavior Modification on Outcome in Early- to Moderate-Stage Chronic Kidney Disease: A Cluster-Randomized Trial. PloS one 11(3):e0151422. [PubMed: 26999730]

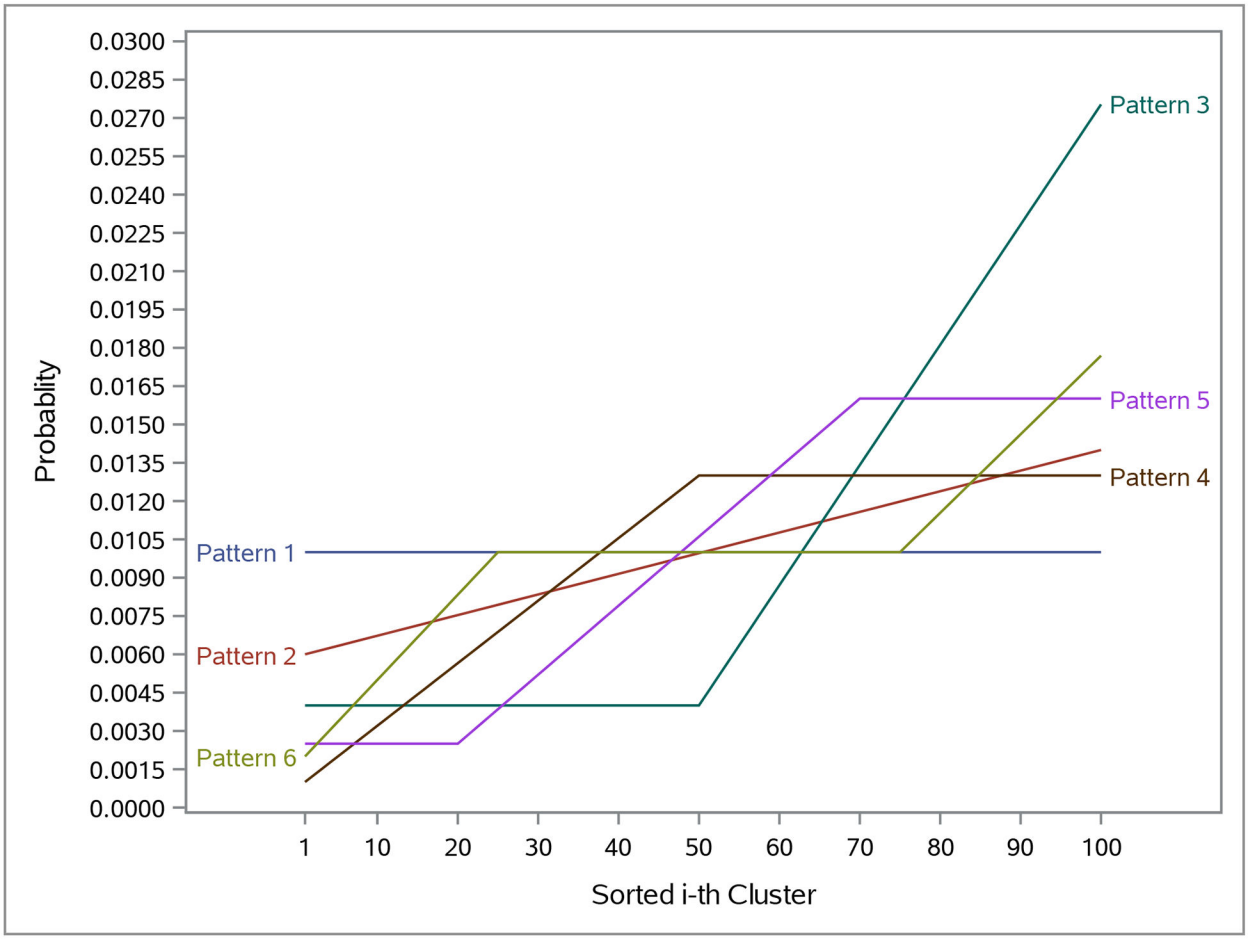

Figure 1.
Six basic patterns of probabilities ($p_1, \cdots, p_m$).

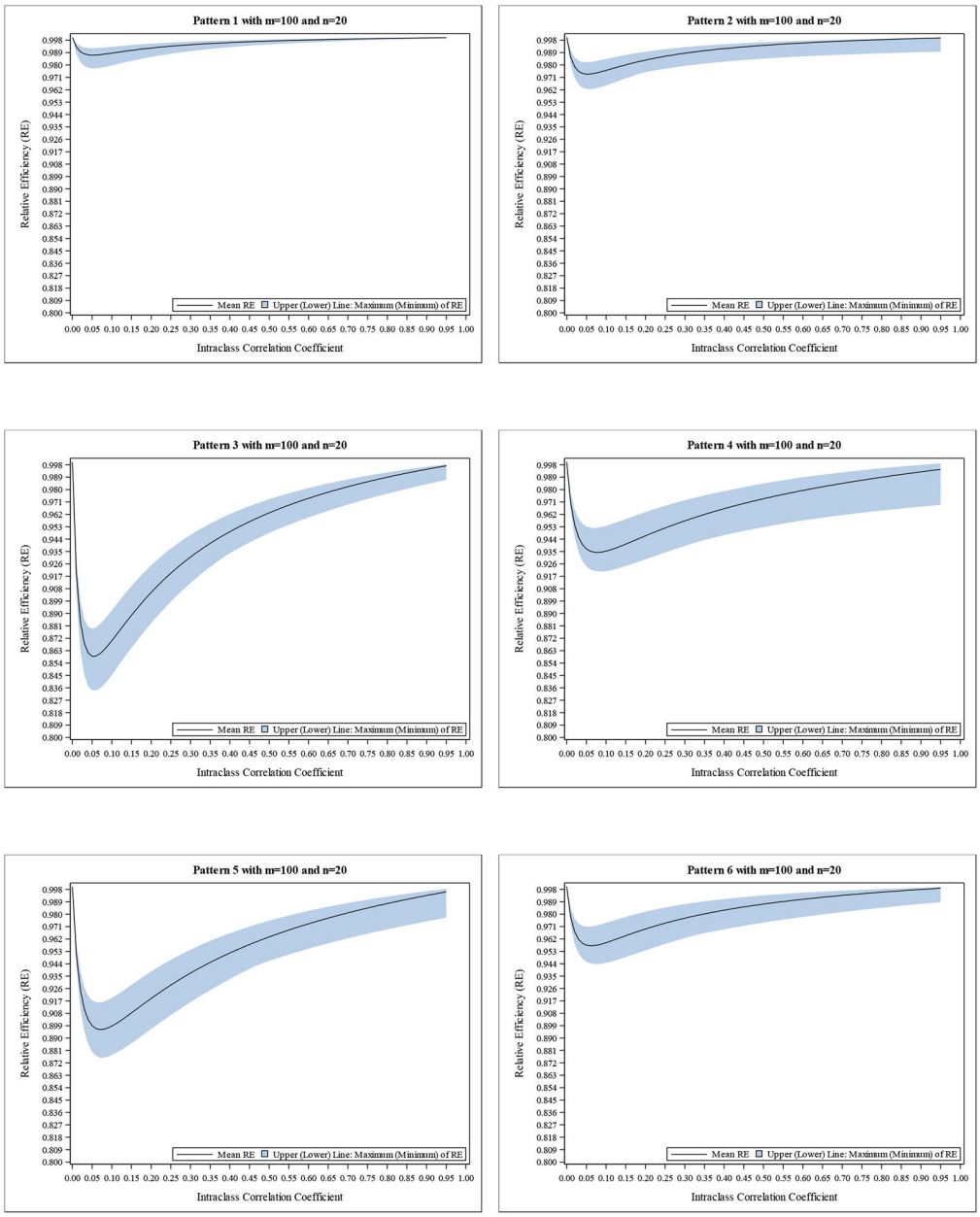

Figure 2.
Relative efficiency of the treatment effect for six patterns, with $m = 100$ and $n = 20$.

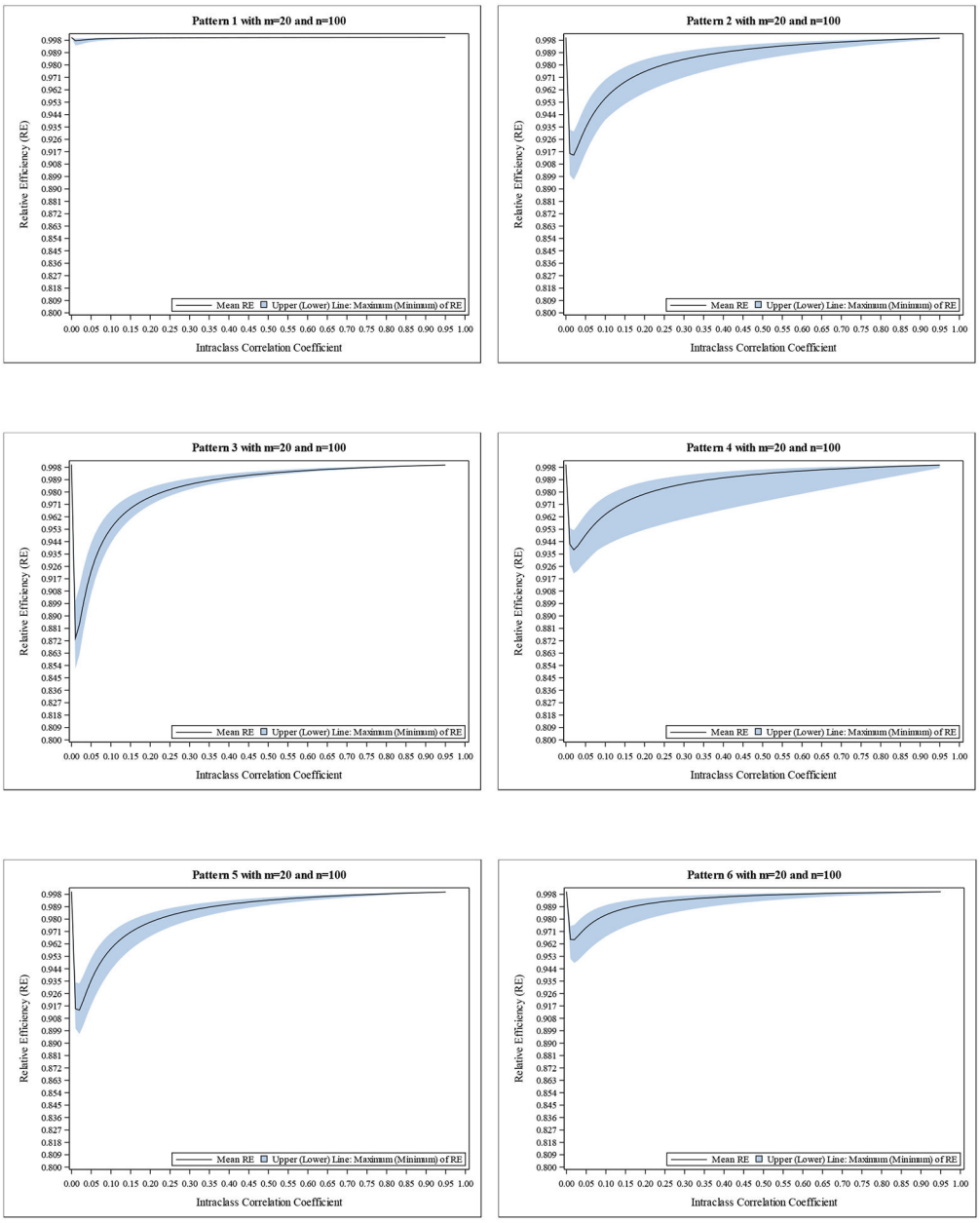

Figure 3.
Relative efficiency of the treatment effect for six patterns, with $m = 20$ and $n = 100$.

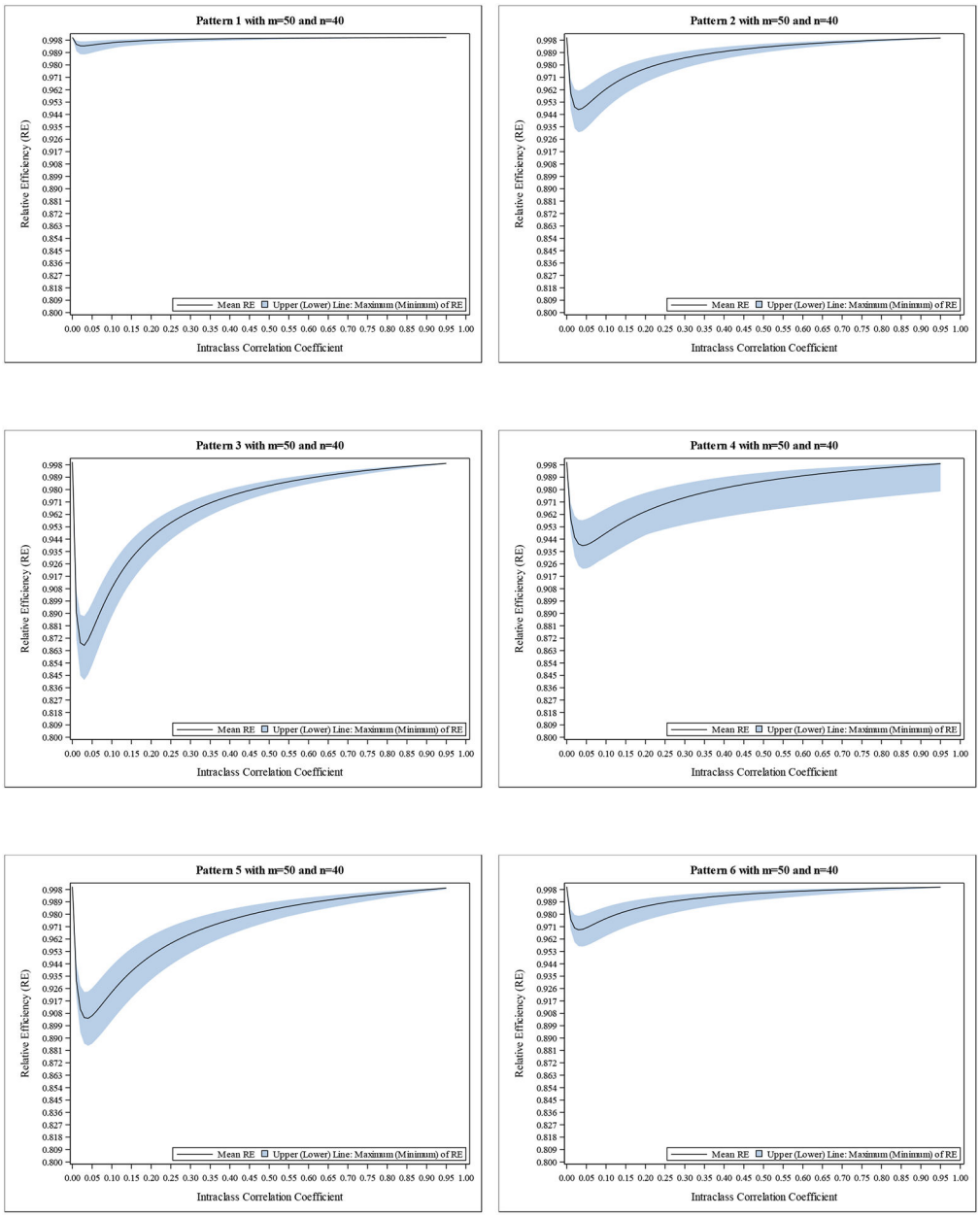

Figure 4.
Relative efficiency of the treatment effect for six patterns, with $m = 50$ and $n = 40$.

Table 1

Multinomial distribution with probabilities $(p_1, \cdots, p_m)$ with $m = 100$ and $n = 20$

Pattern	Description	$(p_1, \cdots, p_m)$[1]	Other Parameters
1	Constant $P_1 = p_2 \cdots = p_m$	$P_i = 1/m$, $i = 1, \cdots, m$	
2	Monotonically increasing $P_1 < p_2 < \cdots < p_m$	${P_1}^{\#} = 0.006$, $p_i = p_1 + (i-1)d$, $i = 2, \cdots, m$	$d = 2\frac{1/m - p_1}{m-1}$
3	Constant followed by monotonically increasing $P_1 = \cdots = p_k < p_{k+1} < \cdots < p_m$	${k}^{\#} = 50$, $p_i = \frac{P_k}{k}, \quad i = 1, \quad \cdots, \quad k$ $p_i = p_k + (i-k)d$, $i = k+1, \cdots, m$	${P_k}^{\#} = \sum_{i=1}^{k} p_i = 0.2$, $d = \frac{2}{m-k}\left(\frac{1 - P_k + \frac{P_k}{k}}{m-k+1} - \frac{P_k}{k}\right)$
4	Monotonically increasing followed by constant $P_1 < \cdots < p_k = p_{k+1} = \cdots = p_m$	${k}^{\#} = 50$ ${P_k}^{\#} = \sum_{i=k+1}^{m} p_i = 0.65$ $p_i = p_k - (k-i)d$, $i = 1, \cdots, k-1$	${P_k}^{\#} = \sum_{i=k+1}^{m} p_i = 0.65$ $d = \frac{2}{k-1}\left(\frac{P_k}{m-k} - \frac{1-P_k}{k}\right)$
5	Constant, monotonically increasing followed by constant $p_1 = p_2 \cdots = p_{k_1} < p_{k_1+1} < \cdots < p_{k_2} = \cdots = p_m$	${k_1}^{\#} = 20$, ${k_2}^{\#} = 70$, $p_i = \frac{P_{k_1}}{k_1}, \quad i = 1, \quad \cdots, \quad k_1$ $p_i = p_{k_1} + (i-k_1)d, i = k_1+1, \quad \cdots, \quad k_2$	${P_{k_1}}^{\#} = \sum_{i=1}^{k_1} p_i = 0.05$, $d = \frac{2\left(1 - mP_{k_1}/k_1\right)}{\left(k_2 - k_1\right)\left[2*\left(m - k_2\right) + \left(k_2 - k_1 + 1\right)\right]}$
6	Monotonically increasing, constant followed by monotonically increasing $p_1 < p_2 \cdots < p_{k_1} = p_{k_1+1} = \cdots = p_{k_2} < \cdots < p_m$	${k_1}^{\#} = 25$, ${k_2}^{\#} = 75$ $p_i = \frac{P_{k_2}}{k_2 - k_1}, \quad i = k_1, \quad \cdots, \quad k_2$ $p_i = p_{k_1} - (k_1 - i)d_1, i = 1, \quad \cdots, \quad k_1 - 1$ $p_i = p_{k_2} + (i-k_2)d_2, i = k_2+1, \quad \cdots, \quad m$ Note: $\frac{k_1}{2}\frac{P_{k_2}}{k_2-k_1} < P_{k_1} < k_1\frac{P_{k_2}}{k_2-k_1}$	${P_{k_1}}^{\#} = \sum_{i=1}^{k_1} p_i = 0.15$, ${P_{k_2}}^{\#} = \sum_{i=k_1+1}^{k_2} p_i = 0.5$ $d_1 = \frac{2}{k_1 - 1}\left(\frac{P_{k_2}}{k_2 - k_1} - \frac{P_{k_1}}{k_1}\right)$, $d_2 = \frac{2}{m-k_2}\left(\frac{1 - P_{k_1} - P_{k_2} + \frac{P_{k_1}}{k_1}}{m - k_2 + 1} - \frac{P_{k_2}}{k_2 - k_1}\right)$

[1] $(p_1, \cdots, p_m)$ are probabilities and $\sum_{i=1}^{m} p_i = 1$.

#: required parameters.

Table 2

Required parameters in simulation designs

Pattern	Total Sample Size (N = 200)			Total Sample Size (N = 600)			Total Sample Size (N = 2000)		
	m = 5	m = 20	m = 40	m = 6	m = 30	m = 60	m = 20	m = 50	m = 100
	n = 40	n = 10	n = 5	n = 100	n = 20	n = 10	n = 100	n = 40	n = 20
2	$p_1 = 0.1$	$p_1 = 0.006$	$p_1 = 0.006$	$p_1 = 0.006$	$p_1 = 0.006$	$p_1 = 0.006$	$p_1 = 0.006$	$p_1 = 0.006$	$p_1 = 0.006$
3	$k = 3$	$k = 10$	$k = 20$	$k = 3$	$k = 15$	$k = 30$	$k = 10$	$k = 25$	$k = 50$
	$p_k = 0.5$	$p_k = 0.4$	$p_k = 0.4$	$p_k = 0.2$	$p_k = 0.2$	$p_k = 0.2$	$p_k = 0.2$	$p_k = 0.2$	$p_k = 0.2$
4	$k = 2$	$k = 10$	$k = 20$	$k = 3$	$k = 15$	$k = 30$	$k = 10$	$k = 25$	$k = 50$
	$p_k = 0.65$	$p_k = 0.65$	$p_k = 0.65$	$p_k = 0.65$	$p_k = 0.65$	$p_k = 0.65$	$p_k = 0.65$	$p_k = 0.65$	$p_k = 0.65$
5	$k_1 = 2$	$k_1 = 4$	$k_1 = 8$	$k_1 = 2$	$k_1 = 6$	$k_1 = 12$	$k_1 = 4$	$k_1 = 10$	$k_1 = 20$
	$k_2 = 3$	$k_2 = 14$	$k_2 = 28$	$k_2 = 4$	$k_2 = 21$	$k_2 = 42$	$k_2 = 14$	$k_2 = 35$	$k_2 = 70$
	$P_{k_1} = 0.05$	$P_{k_1} = 0.05$	$P_{k_1} = 0.05$	$P_{k_1} = 0.05$	$P_{k_1} = 0.05$	$P_{k_1} = 0.05$	$P_{k_1} = 0.05$	$P_{k_1} = 0.05$	$P_{k_1} = 0.05$
6	$k_1 = 2$	$k_1 = 5$	$k_1 = 10$	$k_1 = 2$	$k_1 = 7$	$k_1 = 15$	$k_1 = 5$	$k_1 = 12$	$k_1 = 25$
	$k_2 = 3$	$k_2 = 15$	$k_2 = 30$	$k_2 = 5$	$k_2 = 22$	$k_2 = 45$	$k_2 = 15$	$k_2 = 37$	$k_2 = 75$
	$P_{k_1} = 0.3$	$P_{k_1} = 0.15$	$P_{k_1} = 0.15$	$P_{k_1} = 0.2$	$P_{k_1} = 0.15$	$P_{k_1} = 0.15$	$P_{k_1} = 0.15$	$P_{k_1} = 0.15$	$P_{k_1} = 0.15$
	$P_{k_2} = 0.2$	$P_{k_2} = 0.5$	$P_{k_2} = 0.5$	$P_{k_2} = 0.5$	$P_{k_2} = 0.5$	$P_{k_2} = 0.5$	$P_{k_2} = 0.5$	$P_{k_2} = 0.5$	$P_{k_2} = 0.5$

Table 3

Minimum and median of mean RE for the different simulation study designs

Total Sample Size (*N*)	Number of Clusters (*m*)	Cluster Sizes (*n*)	Pattern	CV Mean (min, max)	Mean RE[1]	
					ρ, Minimum[2]	ρ, Median[2]
200	5	40	1	0.15 (0.02, 0.34)	0.03, 0.9948	0.48–0.49, 0.9994
			2	0.42 (0.17, 0.69)	0.03, 0.9617	0.48–0.49, 0.9952
			3	0.29 (0.06, 0.59)	0.02, 0.9833	0.48–0.49, 0.9983
			4	0.24 (0.06, 0.45)	0.03, 0.9868	0.48–0.49, 0.9984
			5	0.81 (0.69, 0.92)	0.06, 0.7974	0.48–0.49, 0.9311
			6	0.35 (0.13, 0.59)	0.03, 0.9717	0.48–0.49, 0.9962
	20	10	1	0.31 (0.18, 0.47)	0.10, 0.9754	0.49–0.50, 0.9903
			2	0.63 (0.43, 0.84)	0.14, 0.8918	0.51–0.52, 0.9306
			3	0.40 (0.19, 0.61)	0.10, 0.9612	0.49–0.50, 0.9843
			4	0.52 (0.37, 0.68)	0.16, 0.9170	0.52–0.53, 0.9427
			5	0.62 (0.47, 0.81)	0.14, 0.8908	0.50–0.51, 0.9324
			6	0.46 (0.26, 0.66)	0.13, 0.9425	0.50–0.51, 0.9669
	40	5	1	0.44 (0.29, 0.62)	0.21, 0.9482	0.53–0.54, 0.9649
			2	0.63 (0.43, 0.81)	0.24, 0.8924	0.56–0.57, 0.9146
			3	0.51 (0.34, 0.71)	0.21, 0.9338	0.54–0.55, 0.9544
			4	0.60 (0.43, 0.75)	0.28, 0.8945	0.59–0.60, 0.9112
			5	0.70 (0.50, 0.88)	0.28, 0.8600	0.59–0.60, 0.8816
			6	0.55 (0.36, 0.76)	0.24, 0.9175	0.56–0.57, 0.9365
600	6	100	1	0.09 (0.01, 0.19)	0.01, 0.9979	0.48–0.49, 0.9999
			2	0.73 (0.62, 0.83)	0.03, 0.8485	0.48–0.49, 0.9531
			3	0.76 (0.60, 0.91)	0.01, 0.8848	0.48–0.49, 0.9936
			4	0.51 (0.44, 0.60)	0.03, 0.9071	0.48–0.49, 0.9848
			5	0.72 (0.62, 0.84)	0.02, 0.8450	0.48–0.49, 0.9817
			6	0.51 (0.40, 0.64)	0.02, 0.9315	0.48–0.49, 0.9936
	30	20	1	0.22 (0.13, 0.31)	0.05, 0.9876	0.48–0.49, 0.9974

Total Sample Size (N)	Number of Clusters (m)	Cluster Sizes (n)	Pattern	CV Mean (min, max)	Mean RE[1]	
					ρ, Minimum[2]	ρ, Median[2]
2000	60	10	2	0.54 (0.44, 0.65)	0.07, 0.9195	0.48–0.49, 0.9725
			3	0.80 (0.64, 0.92)	0.05, 0.8620	0.48–0.49, 0.9625
			4	0.46 (0.37, 0.55)	0.08, 0.9314	0.49–0.50, 0.9693
			5	0.59 (0.48, 0.69)	0.07, 0.9002	0.48–0.49, 0.9636
			6	0.36 (0.26, 0.46)	0.06, 0.9654	0.48–0.49, 0.9899
			1	0.31 (0.22, 0.44)	0.10, 0.9742	0.49–0.50, 0.9892
			2	0.49 (0.35, 0.61)	0.11, 0.9366	0.49–0.50, 0.9690
			3	0.83 (0.71, 0.96)	0.11, 0.8478	0.49–0.50, 0.9217
			4	0.51 (0.41, 0.58)	0.15, 0.9206	0.50–0.51, 0.9481
			5	0.63 (0.54, 0.72)	0.14, 0.8839	0.51–0.52, 0.9282
			6	0.46 (0.36, 0.56)	0.12, 0.9436	0.50–0.51, 0.9699
	20	100	1	0.09 (0.06, 0.15)	0.01, 0.9976	0.48–0.49, 0.9999
			2	0.56 (0.49, 0.61)	0.02, 0.9146	0.48–0.49, 0.9922
			3	0.77 (0.69, 0.84)	0.01, 0.8738	0.48–0.49, 0.9932
			4	0.43 (0.38, 0.48)	0.02, 0.9380	0.48–0.49, 0.9929
			5	0.55 (0.47, 0.60)	0.02, 0.9140	0.48–0.49, 0.9933
			6	0.35 (0.29, 0.41)	0.02, 0.9651	0.48–0.49, 0.9972
	50	40	1	0.16 (0.11, 0.22)	0.03, 0.9938	0.48–0.49, 0.9993
			2	0.44 (0.38, 0.51)	0.03, 0.9475	0.48–0.49, 0.9925
			3	0.79 (0.72, 0.86)	0.03, 0.8667	0.48–0.49, 0.9819
			4	0.43 (0.37, 0.48)	0.04, 0.9393	0.48–0.49, 0.9855
			5	0.57 (0.51, 0.62)	0.04, 0.9044	0.48–0.49, 0.9820
			6	0.34 (0.28, 0.39)	0.03, 0.9686	0.48–0.49, 0.9952
	100	20	1	0.22 (0.18, 0.29)	0.05, 0.9871	0.48–0.49, 0.9972
			2	0.32 (0.27, 0.38)	0.05, 0.9734	0.48–0.49, 0.9939
			3	0.80 (0.74, 0.88)	0.05, 0.8589	0.48–0.49, 0.9612
			4	0.45 (0.39, 0.49)	0.08, 0.9343	0.48–0.49, 0.9723
			5	0.59 (0.53, 0.64)	0.07, 0.8962	0.48–0.49, 0.9619

Total Sample Size (*N*)	Number of Clusters (*m*)	Cluster Sizes (*n*)	Pattern	CV Mean (min, max)	Mean RE[1]	
					ρ, Minimum[2]	ρ, Median[2]
			6	0.40 (0.34, 0.45)	0.06, 0.9570	0.48–0.49, 0.9867

[1] The mean RE among 1000 simulations is calculated for each ρ.

[2] The minimum and median of mean RE including the corresponding ρ's are identified across all the values of ρ.