Published in final edited form as:

Stat Methods Med Res. 2017 February ; 26(1): 399–413. doi:10.1177/0962280214547381.

Sample size determinations for group-based randomized clinical trials with different levels of data hierarchy between experimental and control arms

Moonseong Heo¹, Alain H. Litwin^{2,3}, Oni Blackstock², Namhee Kim⁴, and Julia H. Arnsten^{1,2,3}

¹Department of Epidemiology and Population Health, Albert Einstein College of Medicine, NY, USA

²Department of Medicine, Montefiore Medical Center, Albert Einstein College of Medicine, NY, USA

³Department of Psychiatry and Behavioral Sciences, Montefiore Medical Center, Albert Einstein College of Medicine, NY, USA

⁴Department of Radiology, Albert Einstein College of Medicine, NY, USA

Abstract

We derived sample size formulae for detecting main effects in group-based randomized clinical trials with different levels of data hierarchy between experimental and control arms. Such designs are necessary when experimental interventions need to be administered to groups of subjects whereas control conditions need to be administered to individual subjects. This type of trial, often referred to as a partially nested or partially clustered design, has been implemented for management of chronic diseases such as diabetes and is beginning to emerge more commonly in wider clinical settings. Depending on the research setting, the level of hierarchy of data structure for the experimental arm can be three or two, whereas that for the control arm is two or one. Such different levels of data hierarchy assume correlation structures of outcomes that are different between arms, regardless of whether research settings require two or three level data structure for the experimental arm. Therefore, the different correlations should be taken into account for statistical modeling and for sample size determinations. To this end, we considered mixed-effects linear models with different correlation structures between experimental and control arms to theoretically derive and empirically validate the sample size formulae with simulation studies.

Keywords

Group-based intervention; sample size; multi-level data; mixed-effects model; varying sizes

Reprints and permissions: sagepub.co.uk/journalsPermissions.nav

Corresponding author: Moonseong Heo, Department of Epidemiology and Population Health, Albert Einstein College of Medicine, 1300 Morris Park Avenue, Belfer 13th floor, Bronx, NY 10461, USA. moonseong.heo@esintein.yu.edu.

Conflict of interest

None declared.

1 Introduction

In clinical trials, interventions are often administered to groups of subjects in an effort to enhance their effects on study outcomes through facilitating social support and reinforcing healthy behaviors among peers.¹ Such group-based treatment models are currently being utilized in clinical care, and are beginning to emerge more commonly.^{2, 3} For example, clinical trials aiming to test the efficacy of group clinical visits (i.e. a group of patients seen simultaneously by a health care provider) have been conducted for the management of diabetes,^{4–6} hepatitis C,⁷ smoking cessation,⁸ and other medical or psychological conditions.^{9–12} Despite such increasing adoption and implementation of group-based interventions in clinical settings, development of rigorous methods to assess statistical power or to determine sample sizes for such approaches has been lacking.

Group-based interventions are usually compared to control interventions that are administered at the individual level of ungrouped participants. This unique aspect imposes a challenge for statistical power assessment or sample size determinations since the hierarchies of data structures between experimental and control arms are different unlike the case of conventional trials which typically assume identical data structures between arms. For example, when groups are randomly selected for the experimental arm and ungrouped individual subjects are randomly selected for the control arm, the levels of hierarchy for the experimental and control arms will be two and one, respectively. This type of design is often referred to as a partially nested or partially clustered design. In such studies, the differences in correlations should be taken into account for both statistical modeling and sample size determinations as illustrated by Roberts and Roberts.¹³ For partially clustered designs under two-level data structures, Bauer et al.¹⁴ discussed several approaches and Baldwin et al.¹⁵ evaluated those statistical models in terms of bias of variance components, type I error and power with extensive simulations, yet without theoretical derivations. Moerbeek and Wong¹⁶ theoretically derived sample size formulae for both continuous and binary outcome data with equal group sizes but did not validate these with simulation studies.

In this study, we extend the partially clustered design to larger scale trials that use multiple centers in which groups and individual subjects are nested. Thus, the levels of hierarchy for the experimental and control arms will be three and two, respectively. The aim of this paper was to derive power functions for testing main effects of group-based interventions in comparison to individual-based control conditions based on two- and three-level mixed-effects linear models. Both equal and varying cluster sizes are considered, and simulation studies are conducted for validation of derived sample formulae. Throughout this paper, the continuous outcome will be denoted by Y , and the arm indicator will be denoted by $X=0$ for control arm and $=1$ for experimental arm. The number of nested units for each level will vary across nesting units. The sets of indices indicating observations from the experimental and control arm subjects will be denoted by E and C , respectively. Although the nomenclatures for units of levels should depend on study context, here we refer to “center”, “group”, and “subject” as the third, the second, and the first level data units, respectively.

2 Two-level model

2.1 Statistical model

When the group-based experimental arm assumes a two level data structure, and the control arm assumes a one level data structure, a pertinent mixed-effects linear model can be formulated as follows

$$Y_{jk}^{(2)} = \beta_0 + \delta^{(2)} X_{jk} + u_j X_{jk} + e_{jk} \quad (1)$$

The two sets, E and C, are defined as: $E = \{j, k \mid X_{jk} = 1\}$ and $C = \{j, k \mid X_{jk} = 0\}$. Groups in the experimental arm are indexed by $j = 1, 2, \dots, J$ for $j \in E$ so that $\#\{j \mid j \in E\} = J$, where $\#\{\cdot\}$ denotes the number of elements in the set $\{\cdot\}$. Although there are no groups in the control arm, we assign “pseudo” group indices for observations from the control arm subjects as follows: $j = J + 1, J + 2, \dots, J + J'$ for $j \in C$ so that $\#\{j \mid j \in C\} = J'$. Subjects are indexed by $k = 1, 2, \dots, K_j$ and $K_j > 1$ for $j \in E$ and $K_j = 1$ for $j \in C$ so that each subject serves as his/her own group in the control arm. The total number of subjects in the

experimental arm is denoted by $N_E = \sum_{j=1}^J K_j$, whereas that in the control arm is denoted by $N_C = \sum_{j=J+1}^{J+J'} K_j = J'$.

The fixed-effect parameters β_0 and $\delta^{(2)}$ represent the intercept and the main effect of the experimental intervention on the outcome $Y^{(2)}$, respectively. Group-specific random effects in the experimental arm are denoted by $u_j \sim N(0, \sigma_u^2)$ for $j \in E$. The random noise is denoted by e_{jk} which is assumed to be $e_{jk} \sim N(0, \sigma_e^2)$. Furthermore, although u_j and e_{jk} are assumed to be mutually independent (mutual independence assumption), the elements of e_{jk} are assumed to be independent over k for given u_j (conditional independence assumption). We assume that the magnitudes of all variance components in model (1) are known.

Under these assumptions, it can be shown that $E(Y_{jk}^{(2)}) = \beta_0 + \delta^{(2)} X_{jk}$ and

$Var(Y_{jk}^{(2)}) = \sigma_e^2 + \sigma_u^2 X_{jk}$. Let $\sigma^2 \equiv Var(Y_{jk}^{(2)} \mid X_{jk} = 1) = \sigma_e^2 + \sigma_u^2$ denote the variance of $Y^{(2)}$ in the experimental arm, then the intra-class correlation coefficient (ICC) of outcome Y within groups in the experimental arm can be expressed as

$$\rho \equiv Corr(Y_{jk}, Y_{jk'}) = \sigma_u^2 / \sigma^2 \quad (2)$$

for $k \neq k'$ and $j \in E$. The variance of $Y^{(2)}$ in the control arm is simply equal to σ_e^2 , i.e.

$Var(Y_{jk}^{(2)} \mid X_{jk} = 0) = \sigma_e^2 = (1 - \rho)\sigma^2 \leq \sigma^2$ for $j \in C$. The null hypothesis for testing the significance of the main effect of experimental intervention is $H_0: \delta^{(2)} = 0$.

2.2 Parameter estimates and variances

An estimate of $\delta^{(2)}$ under model (1) can be obtained as

$$\hat{\delta}^{(2)} = \bar{Y}_E^{(2)} - \bar{Y}_C^{(2)}$$

where

$$\bar{Y}_E^{(2)} = \sum_{j=1}^J \sum_{k=1}^{K_j} Y_{jk} / \sum_{j=1}^J K_j = \sum_{j=1}^J \sum_{k=1}^{K_j} Y_{jk} / N_E \quad \text{and} \quad \bar{Y}_C^{(2)} = \sum_{j=J+1}^{J+J'} Y_{jk} / \sum_{j=J+1}^{J+J'} K_j = \sum_{j=J+1}^{J+J'} Y_{jk} / N_C$$

are the sample means of $Y^{(2)}$ in the experimental and control arms, respectively. The variances of these means can be expressed as

$$Var(\bar{Y}_E^{(2)}) = \sigma^2 / N_E + \left(\sum_{j=1}^J K_j^2 / N_E^2 - 1 / N_E \right) \sigma^2 = \sigma^2 \left\{ (1-\rho) / N_E + \rho \sum_{j=1}^J K_j^2 / N_E^2 \right\}$$

and

$$Var(\bar{Y}_C^{(2)}) = \sigma_e^2 / N_C = (1-\rho) \sigma^2 / N_C.$$

It follows that

$$Var(\hat{\delta}^{(2)}) = Var(\bar{Y}_E^{(2)}) + Var(\bar{Y}_C^{(2)}) = \sigma^2 \left\{ (1-\rho)(1/N_E + 1/N_C) + \rho \sum_{j=1}^J K_j^2 / N_E^2 \right\} \quad (3)$$

2.3 Test statistics, power functions and sample size formulae

The following test statistic $D^{(2)}$ can be used to test the null hypothesis $H_0: \delta^{(2)} = 0$

$$D^{(2)} = \frac{\hat{\delta}^{(2)}}{\sqrt{Var(\hat{\delta}^{(2)})}} = \frac{\bar{Y}_E^{(2)} - \bar{Y}_C^{(2)}}{\sigma \sqrt{(1-\rho)(1/N_E + 1/N_C) + \rho \sum_{j=1}^J K_j^2 / N_E^2}}.$$

The power function $D^{(2)}$ at a two-sided significance level of α can be expressed as follow

$$\varphi^{(2)} \equiv \varphi(D^{(2)}) = \Phi \left\{ \Delta_{(2)} / \sqrt{(1-\rho)(1/N_E + 1/N_C) + \rho \sum_{j=1}^J K_j^2 / N_E^2} - \Phi^{-1}(1-\alpha/2) \right\} \quad (4)$$

where $\Delta_{(2)} = |\delta^{(2)}|/\sigma$ is a standardized effect size or Cohen's d^{17} and Φ is the cumulative distribution function (CDF) of a standard normal distribution. Determination of varying group sizes K_j 's, or N_E , for fixed N_C can be made by solving equation (4) iteratively for K_j 's. On the other hand, determination of sample size N_C for fixed group sizes K_j 's, or N_E should be straightforward by solving equation (4) for N_C .

For special cases of equal number of units, sample size determinations are much more tractable. Suppose that group sizes are equal in the experimental arm, i.e. $K_j = K$ for all $j \in E$ so that $N_E = JK$, and that total numbers of subjects are the same between the two arms, i.e. $N_E = JK = N_C$, then the variance (3) can be reduced to

$$Var \left(\hat{\delta}_E^{(2)} \right) = \frac{\sigma^2}{JK} \{2 + \rho(K-2)\}$$

which simplifies the power function (4) to

$$\varphi^{(2)} = \Phi \left\{ \Delta_{(2)} \sqrt{\frac{JK}{2 + \rho(K-2)}} - \Phi^{-1}(1-\alpha/2) \right\} \quad (5)$$

It follows that the number of groups for fixed group sizes in E can be expressed as

$$J = \frac{\{2 + \rho(K-2)\} z_{\alpha,\varphi}^2}{K \Delta_{(2)}^2} \quad (6)$$

and the group sizes for fixed number of groups in E can be expressed as

$$K = \frac{2(1-\rho) z_{\alpha,\varphi}^2}{J \Delta_{(2)}^2 - \rho z_{\alpha,\varphi}^2} \quad (7)$$

for $\rho < J \Delta_{(2)}^2 / z_{\alpha,\varphi}^2$, where

$$z_{\alpha,\varphi} = \Phi^{-1}(1-\alpha/2) + \Phi^{-1}(\varphi) \quad (8)$$

and Φ^{-1} is the inverse CDF of a standard normal distribution. We note that when K needs to be determined for a desired level of power for a given J , it is possible that K cannot be determined, especially when J is small and ρ (2) is large, resulting in a combination which can make the denominator of equation (7) negative. Therefore, when only a limited number of groups J are feasible to form, the correlation ρ in particular must be small enough for $\rho < J\Delta_{(2)}^2 / z_{\alpha,\varphi}^2$ to be true in equation (7).

3 Three-level model

3.1 Statistical model

When the group-based intervention arm assumes a three level data structure while the control arm assumes a two level data structure, model (1) can be extended as follow

$$Y_{ijk}^{(3)} = \beta_0 + \delta^{(3)} X_{ijk} + u_i + u_{j(i)} X_{ijk} + e_{ijk} \quad (9)$$

The two sets are defined as $E = \{i, j, k \mid X_{ijk} = 1\}$ and $C = \{i, j, k \mid X_{ijk} = 0\}$. Centers are indexed by $i = 1, 2, \dots, I$ for $i \in E$ so that $\#\{i \mid i \in E\} = I$ and $i = I+1, I+2, \dots, I+I'$ for $i \in C$ so that $\#\{i \mid i \in C\} = I'$. Groups are indexed by $j = 1, 2, \dots, J_i$ and subjects by $k = 1, 2, \dots, K_{ij}$. Concerning the number of groups within centers, $J_i > 1$ for $i \in E$, and $J_i = 1$ for $i \in C$, we assign pseudo group indices so that each center serves as its own single nesting pseudo group in the control arm. Similarly, concerning group sizes (or the number of subjects within groups), $K_{ij} = K_i$ for $i \in C$. The total number of subjects in E is denoted by

$$N_E = \sum_{i=1}^I \sum_{j=1}^{J_i} K_{ij} \text{ whereas that in } C \text{ is denoted by } N_C = \sum_{i=I+1}^{I+I'} K_{ij} = \sum_{i=I+1}^{I+I'} K_i.$$

The fixed-effect parameters β_0 and $\delta^{(3)}$ represent the intercept and the main effect of the experimental intervention on the outcome $Y^{(3)}$, respectively, whereas the center-specific random intercepts are denoted by $u_i \sim N(0, \sigma_3^2)$. The group-specific random intercepts within the experimental arm centers are denoted by $u_{j(i)} \sim N(0, \sigma_2^2)$ for $j \in E$. The random noise is denoted by e_{ijk} which is assumed to be $e_{ijk} \sim N(0, \sigma_e^2)$. These three random components are assumed to be mutually independent (mutual independence assumption). However, the elements of e_{ijk} are assumed to be independent over k for given u_i and u_j and those of u_j are independent over j for given u_i (conditional independence assumption). Again, we assume that the magnitudes of all variance components in model (9) are known.

It follows that $E(Y_{ijk}^{(3)}) = \beta_0 + \delta^{(3)} X_{ijk}$ and $Var(Y_{ijk}^{(3)}) = \sigma_e^2 + \sigma_2^2 X_{ijk} + \sigma_3^2$. Let $\sigma^2 \equiv Var(Y_{ijk}^{(3)} \mid X_{ijk} = 1) = \sigma_e^2 + \sigma_2^2 + \sigma_3^2$ denote the variance of $Y^{(3)}$ in the experimental arm. Then, the correlations among the group-level observations can be obtained as

$$\rho_2 \equiv Corr(Y_{ijk}, Y_{ij'k'}) = \sigma_3^2 / \sigma^2 \quad (10)$$

for $j \neq j'$ and $i \in E$. The correlations among the subject-level observations can be obtained as

$$\rho_1 \equiv \text{Corr}(Y_{ijk}, Y_{ijk'}) = (\sigma_2^2 + \sigma_3^2) / \sigma^2 \quad (11)$$

for $k \neq k'$ and $i \in E$. On the other hand, the variance of $Y^{(3)}$ in the control arm is denoted by $\text{Var}(Y_{ijk}^{(3)} | X_{ijk}=0) = \sigma_e^2 + \sigma_3^2 = (1 - \rho_1 + \rho_2) \sigma^2 \leq \sigma^2$ since $\rho_2 \leq \rho_1$. The null hypothesis for testing the significance of the main effect of experimental intervention is $H_0: \delta^{(3)} = 0$.

3.2 Parameter estimates and variances

An estimate of $\delta^{(3)}$ under model (9) can be obtained as

$$\hat{\delta}^{(3)} = \bar{Y}_E^{(3)} - \bar{Y}_C^{(3)}$$

where

$$\begin{aligned} \bar{Y}_E^{(3)} &= \sum_{i=1}^I \sum_{j=1}^{J_i} \sum_{k=1}^{K_{ij}} Y_{ijk} / \sum_{i=1}^I \sum_{j=1}^{J_i} K_{ij} = \sum_{i=1}^I \sum_{j=1}^{J_i} \sum_{k=1}^{K_{ij}} Y_{ijk} / N_E \quad \text{and} \\ \bar{Y}_C^{(3)} &= \sum_{i=I+1}^{I+I'} \sum_{k=1}^{K_{ij}} Y_{ijk} / \sum_{i=I+1}^{I+I'} K_i = \sum_{i=I+1}^{I+I'} \sum_{k=1}^{K_{ij}} Y_{ijk} / N_C \end{aligned}$$

are the sample means of $Y^{(3)}$ in the experimental and control arms, respectively. The variances of these means can be expressed as

$$\text{Var}(\bar{Y}_E^{(3)}) = \sigma^2 \left\{ \frac{1 - \rho_1}{N_E} + \frac{1}{N_E^2} \left(\rho_1 \sum_{i=1}^I \sum_{j=1}^{J_i} K_{ij}^2 + 2\rho_2 \sum_{i=1}^I \sum_{j=1}^{J_i} \sum_{j'=1}^{J_i} K_{ij} K_{ij'} \right) \right\} \quad (12)$$

and

$$\text{Var}(\bar{Y}_C^{(3)}) = \sigma^2 \left\{ (1 - \rho_1) / N_C + \rho_2 \sum_{i=I+1}^{I+I'} K_i^2 / N_C^2 \right\} \quad (13)$$

It follows that

$$\begin{aligned}
 \text{Var} \left(\hat{\delta}_E^{(3)} \right) &= \text{Var} \left(\bar{Y}_E^{(3)} \right) + \text{Var} \left(\bar{Y}_C^{(3)} \right) \\
 &= \sigma^2 \left\{ (1-\rho_1) \left(\frac{1}{N_E} + \frac{1}{N_C} \right) + \rho_2 \left(\frac{2}{N_E^2} \sum_{i=1}^I \sum_{j=1}^{J_i} \sum_{j' > j}^{J_i} K_{ij} K_{ij'} + \frac{1}{N_C^2} \sum_{i=I+1}^{I+I'} K_i^2 \right) + \frac{\rho_1}{N_E^2} \sum_{i=1}^I \sum_{j=1}^{J_i} K_{ij}^2 \right\} \quad (14)
 \end{aligned}$$

3.3 Test statistics, power functions and sample size formulae

The following test statistic $D^{(3)}$ can be used to test the null hypothesis $H_0: \delta^{(3)} = 0$

$$\begin{aligned}
 D^{(3)} &= \frac{\hat{\delta}^{(3)}}{\sqrt{\text{Var}(\hat{\delta}^{(3)})}} \\
 &= \frac{\bar{Y}_E^{(3)} - \bar{Y}_C^{(3)}}{\sigma \sqrt{(1-\rho_1) \left(\frac{1}{N_E} + \frac{1}{N_C} \right) + \rho_2 \left(\frac{2}{N_E^2} \sum_{i=1}^I \sum_{j=1}^{J_i} \sum_{j' > j}^{J_i} K_{ij} K_{ij'} + \frac{1}{N_C^2} \sum_{i=I+1}^{I+I'} K_i^2 \right) + \frac{\rho_1}{N_E^2} \sum_{i=1}^I \sum_{j=1}^{J_i} K_{ij}^2}}.
 \end{aligned}$$

The power function of $D^{(3)}$ at a two-sided significance level of α can be expressed as follow

$$\begin{aligned}
 \varphi^{(3)} &\equiv \varphi(D^{(3)}) \\
 &= \Phi \left\{ \Delta_{(3)} / \sqrt{(1-\rho_1) \left(\frac{1}{N_E} + \frac{1}{N_C} \right) + \rho_2 \left(\frac{2}{N_E^2} \sum_{i=1}^I \sum_{j=1}^{J_i} \sum_{j' > j}^{J_i} K_{ij} K_{ij'} + \frac{1}{N_C^2} \sum_{i=I+1}^{I+I'} K_i^2 \right) + \frac{\rho_1}{N_E^2} \sum_{i=1}^I \sum_{j=1}^{J_i} K_{ij}^2} - \Phi^{-1}(1-\alpha/2) \right\} \quad (15)
 \end{aligned}$$

where $\Delta_{(3)} = |\delta^{(3)}|/\sigma$ is again a standardized effect size. The determinations of varying group sizes in the experimental arm and cluster sizes in both arms should be made in an iterative manner.

For special cases where there are equal number of units, sample size determinations are much more tractable. If the sizes of units are equal for both arms, i.e. $J_E = J$ for all $i \in E$ and $K_E = K$ for all $j \in E$ so that $N_E = IJK$ in the experimental arm and $K_C = K$ for all $i \in C$ and $I' = I$ so that $N_C = N_E = IJK$ in the control arm (total numbers of subjects are the same between the two arms), then the variances (12), (13) and (14) can be, respectively, reduced to

$$\begin{aligned}
 \text{Var} \left(\bar{Y}_E^{(3)} \right) &= \frac{\sigma^2}{IJK} \{1 + (K-1)\rho_1 + K(J-1)\rho_2\}, \\
 \text{Var} \left(\bar{Y}_C^{(3)} \right) &= \frac{\sigma^2}{IJK} (1 - \rho_1 + JK\rho_2)
 \end{aligned}$$

and

$$Var\left(\hat{\delta}_{\text{E}}^{(3)}\right)=\frac{\sigma^2}{IJK}\{2+(K-2)\rho_1+K(2J-1)\rho_2\}.$$

It follows that the power function (15) can be simplified to

$$\varphi^{(3)} \equiv \varphi(D^{(3)}) = \Phi\left\{\Delta_{(3)}\sqrt{IJK/\{2+(K-2)\rho_1+K(2J-1)\rho_2\}} - \Phi^{-1}(1-\alpha/2)\right\} \quad (16)$$

Subsequently, the sample size of each level can be determined as follow

$$I = \frac{\{2+(K-2)\rho_1+K(2J-1)\rho_2\} z_{\alpha,\varphi}^2}{JK\Delta_{(3)}^2} \quad (17)$$

$$J = \frac{\{2+(K-2)\rho_1-K\rho_2\} z_{\alpha,\varphi}^2}{IK\Delta_{(3)}^2 - 2K\rho_2 z_{\alpha,\varphi}^2} \quad (18)$$

for $\rho_2 < IK\Delta_{(3)}^2/2Kz_{\alpha,\varphi}^2$, and

$$K = \frac{2(1-\rho_1)z_{\alpha,\varphi}^2}{IJ\Delta_{(3)}^2 - \{\rho_1+(2J-1)\rho_2\} z_{\alpha,\varphi}^2} \quad (19)$$

for combinations ρ_1 (11) and ρ_2 (10) that makes the denominator of equation (19) positive, where $z_{\alpha,\varphi}$ is defined as in (8). Again, we note that determinations of J (18) with given I and K or determinations of K (19) with given I and J may not be possible for desired power with certain combinations of ρ_1, ρ_2 , and IK or IJ , respectively. For example, ρ_2 in particular must be very small for small ρ_1, I and J in equation (19).

4 Simulation studies and results

We conducted simulation studies to validate the derived sample size formulae for both two-level models (1) and three-level models (9). We used SAS PROC MIXED to fit those models with unknown variances which are usually assumed in practice, and to compute simulation-based empirical power. This computation was based on critical values of t -distributions under the null hypotheses with degrees of freedom determined based on the method proposed by Kenward and Roger.¹⁸ Throughout this section, a nominal statistical power is set at 80%, a two-sided significance level is set at $\alpha=0.05$, and 1000 simulations were

conducted for each combination of design parameters. Therefore, the 95% confidence intervals (CI) for the empirical power will be 0.8 ± 0.025 .

4.1 Two-level model

4.1.1 Equal group sizes—We first determined J , the number of groups in the experimental arm in equation (6), for 80% statistical power with other given parameters including group size K , the number of subjects per group. Accordingly the sample size, or the number of subjects, the control arm is determined as $N_C = JK$. We then computed empirical power, denoted by $\tilde{\phi}^{(2)}$, by fitting model (1) based on 1000 simulations for each combination of the specified parameters in Table 1. The results in Table 1 show that the theoretical power $\phi^{(2)}$ (5) and the simulation-based empirical power $\tilde{\phi}^{(2)}$ is very close with $\text{mean}(\phi^{(2)}) - \text{mean}(\tilde{\phi}^{(2)}) = 0.014$ (or 1.8% bias) and $\max|\phi^{(2)} - \tilde{\phi}^{(2)}| = 0.034$, which is tolerable compared to the 0.025 margin of the 95% CI. This finding supports the validity of the power function $\phi^{(2)}$ (5) and sample size formulae for J in equation (6).

Second, we determined K , the number of subjects per group in the experimental arm in equation (7), for 80% statistical power with other given parameters including group size J . In this case, we considered small J to evaluate the power function under this condition. The other parameters are specified in Table 2, and especially the correlation $\rho^{(2)}$ had to be extremely small to ensure a positive K in equation (7) as noted earlier. The results in Table 2 show that $\phi^{(2)}$ and $\tilde{\phi}^{(2)}$ are still close with $\text{mean}(\phi^{(2)}) - \text{mean}(\tilde{\phi}^{(2)}) = 0.054$ (or 6.8% bias) and $\max|\phi^{(2)} - \tilde{\phi}^{(2)}| = 0.075$, which is large compared to the 0.025 margin of the 95% CI. Furthermore, the theoretical power $\phi^{(2)}$ is underestimated $\tilde{\phi}^{(2)}$ for all combinations, especially for very large K compared to J . This finding cautions the use of the power function (5) for determining K in equation (7) with small J and very small ρ .

4.1.2 Unequal group sizes—Although the power function (4) should be applied to the cases of unequal group sizes, sample size determinations require iterative solutions as mentioned earlier. Therefore, we assessed approximate applicability of the sample size J (6) for equal unit sizes to the case of unequal group sizes. To this end, we determined J for a given equal group size K and then considered a uniform random variable $U(a, b)$ with expectation K to randomly draw varying group sizes K_j where the integer values of a and b were determined as follows: $a = K - \text{floor}(K/2)$ and $b = K + \text{floor}(K/2)$ so that $a > 0$ and $E\{U(a, b)\} = (a+b)/2 = K$, where the function $\text{floor}(x)$ returns the greatest integer smaller than or equal to x . For the control arm, however, we fixed sample size for practical reasons at $N_C = JK$ which is equal to average sample size for the experimental arm.

The theoretical power $\phi^{(2)}$ was based on the mean group sizes for the experimental group and the fixed sample size for the control group; however, the empirical power $\tilde{\phi}^{(2)}$ was based on the randomly drawn unequal group sizes in the experimental arm. The results in Table 3 again show that the theoretical power and the simulation-based empirical power is very close with $\text{mean}(\phi^{(2)}) - \text{mean}(\tilde{\phi}^{(2)}) = 0.026$ (or 3.3% bias) and $\max|\phi^{(2)} - \tilde{\phi}^{(2)}| = 0.058$, which is somewhat large compared to the 0.025 margin of the 95% CI. Therefore, when group sizes in the experimental arm need to vary, the mean group sizes can be used as K to determine J in equation (6). Sample size for the control group can be determined accordingly as JK .

4.2 Three-level model

4.2.1 Equal unit sizes—We first determined I , the number of centers in the experimental arm in equation (17), for 80% statistical power with other given parameters including number of groups J per center and number of subjects K per group. The sample size, or the number of subjects, per center for the control arm is determined as JK so that $N_E = N_C = IJK$. We then computed empirical power, denoted by $\tilde{\phi}^{(3)}$, by fitting model (9) based on 1000 simulations for each combination of the specified parameters in Table 4. The results in Table 4 show that the theoretical power $\phi^{(3)}$ (16) and the simulation-based empirical power $\tilde{\phi}^{(3)}$ is very close with $\text{mean}(\phi^{(3)}) - \text{mean}(\tilde{\phi}^{(3)}) = 0.006$ (0.8% bias) and $\max|\phi^{(3)} - \tilde{\phi}^{(3)}| = 0.027$, which is excellent compared to the 0.025 margin of the 95% CI. This finding supports the validity of the power function $\phi^{(3)}$ (16) and sample size formulae for I in equation (17).

Second, we determined K , the number of subjects per group in the experimental arm in equation (19), for 80% statistical power with other given parameters including I and J . In this case, again, we considered small I and J to evaluate the power function under this condition. The other parameters are specified in Table 5, and especially the correlation ρ_2 (10) had to be extremely small to ensure a positive K in equation (19) as noted earlier. The results in Table 5 show that $\phi^{(3)}$ and $\tilde{\phi}^{(3)}$ are close with $\text{mean}(\phi^{(3)}) - \text{mean}(\tilde{\phi}^{(3)}) = 0.035$ (or 4.4% bias) and $\max|\phi^{(3)} - \tilde{\phi}^{(3)}| = 0.050$, which is somewhat large compared to the 0.025 margin of the 95% CI. Compared to the case of small J under two-level models, however, the differences between $\phi^{(3)}$ and $\tilde{\phi}^{(3)}$ for all combinations are smaller and tolerable despite very small ρ_2 . This finding supports the use of the power function (16) even for small I and J with very small ρ_2 in determination of K in equation (19).

4.2.2 Unequal unit sizes—Again, although the power function (15) should be applied to the cases of unequal unit sizes, sample size determinations require iterative solutions as mentioned earlier. Therefore, we assessed again approximate applicability of the sample size I (17) for equal unit sizes to the case of unequal unit sizes. To this end, we determined I for an equal number of groups J and an equal group size K and then randomly drew varying number of groups J_i from a uniform distribution $U(a, b)$ with $a = J - \text{floor}(J/2)$ and $b = J + \text{floor}(J/2)$ so that $a > 0$ and $E\{U(a, b)\} = J$. We further varied the group size K_{ij} drawing uniform distribution $U(a, b)$ with $a = K - \text{floor}(K/2)$ and $b = K + \text{floor}(K/2)$ so that $a > 0$ and $E\{U(a, b)\} = K$. Likewise, for the control arm, we varied center size, or number of subjects per center, K_i by randomly drawing from a uniform distribution $U(a, b)$ with $a = JK - \text{floor}(JK/2)$ and $b = JK + \text{floor}(JK/2)$ so that $a > 0$ and $E\{U(a, b)\} = JK$.

The theoretical power $\phi^{(3)}$ was based on the mean unit sizes for the experimental group and the mean center size for the control group; however, the empirical power $\tilde{\phi}^{(3)}$ was based on the randomly drawn unequal unit sizes. The results in Table 6 show that the theoretical power and the simulation-based empirical power is very close with $\text{mean}(\phi^{(3)}) - \text{mean}(\tilde{\phi}^{(3)}) = 0.014$ (1.8% bias) and $\max|\phi^{(3)} - \tilde{\phi}^{(3)}| = 0.030$, which is acceptable compared to the 0.025 margin of the 95% CI. Therefore, when group sizes need to be varied, the mean center sizes and group sizes can be used as J and K , respectively, to determine I in equation (17). Likewise, the center sizes for the control arm can vary with mean JK .

5 Discussion

We derived and validated sample size formulae for detecting main effects for trials with different levels of data hierarchy between arms with varying sizes of units of each nested level. In fact, it can be shown that the sample size $J(6)$ is smaller than that for a trial with two levels of data hierarchy for both arms.¹⁹ Likewise, the sample size $I(17)$ is smaller than that for a trial with three levels of data hierarchy for both arms.^{20,21} Through all Tables from 1 to 6, the mean theoretical power was greater than the mean empirical power (see the bottom rows). This discrepancy was due to the fact that the theoretical power was derived under known variance components, using standard normal distributions, whereas the empirical power was estimated under unknown variance components using t -distributions. The minimal nature of the discrepancies, however, assures that the theoretical power can be used for models with unknown variance components that have to be replaced by estimates for fitting regardless of equal or varying sizes of nesting units. Nevertheless, cautions should be exercised for determining K with small J and very small ρ under two-level partially clustered designs.

These derivations have great relevance to clinical research, since the Affordable Care Act and various health care system stakeholders encourage the development and evaluation of innovative delivery models such as group models of care.²² In addition, as new study designs emerge to expand the field of patient centered outcomes research, it is imperative to evaluate the implications of design innovations and to ensure that design methodology remains rigorous. Studies in which there are systematic, planned differences between comparison groups are likely to be more frequent, and it is crucially important to understand how such differences affect the interpretation of results. Therefore, development of rigorous statistical methods will be needed in order to establish that group-based models of care improve clinical outcomes compared to individual-based models of care. To this end, for future studies, two potential extensions of the sample size determination approaches proposed here would be valuable: (1) extension to designs that implement both group-based experimental and individual-based control arms within centers; and (2) extension to designs with binary outcomes and other types of outcomes.

Acknowledgments

We are grateful to two anonymous reviewers for their corrections and suggestions for improving the contents of this paper.

Funding

This study was supported in part by the following NIH grants: R01DA034086, R25DA023021, K23MH102129, P30AI051519, and UL1TR001073.

References

1. Jaber R, Braksmajer A, Trilling JS. Group visits: A qualitative review of current research. *J Am Board Family Med.* 2006; 19:276–290.
2. Drum D, Becker M, Hess E. Expanding the application of group interventions: emergence of groups in health care settings. *J Specialist Group Work.* 2011; 36:247–263.

3. McCarthy C, Hart S. Designing groups to meet evolving challenges in health care settings. *J Specialist Group Work*. 2011; 36:352–367.
4. Wagner EH, Grothaus LC, Sandhu N, et al. Chronic care clinics for diabetes in primary care – A system-wide randomized trial. *Diab Care*. 2001; 24:695–700.
5. Sadur CN, Moline N, Costa M, et al. Diabetes management in a health maintenance organization – Efficacy of care management using cluster visits. *Diabetes Care*. 1999; 22:2011–2017. [PubMed: 10587835]
6. Burke RE, O’Grady ET. Group visits hold great potential for improving diabetes care and outcomes, but best practices must be developed. *Health Affairs*. 2012; 31:103–109. [PubMed: 22232100]
7. Stein MR, Soloway IJ, Jefferson KS, et al. Concurrent group treatment for hepatitis C: Implementation and outcomes in a methadone maintenance treatment program. *J Substance Abuse Treatment*. 2012; 43:424–432.
8. Moadel AB, Bernstein SL, Mermelstein RJ, et al. A randomized controlled trial of a tailored group smoking cessation intervention for HIV-infected smokers. *JAIDS*. 2012; 61:208–215. [PubMed: 22732470]
9. Ickovics JR, Kershaw TS, Westdahl C, et al. Group prenatal care and perinatal outcomes: A Randomized controlled trial. *Obstet Gynecol*. 2007; 110:937–937.
10. Scott JC, Conner DA, Venohr I, et al. Effectiveness of a group outpatient visit model for chronically ill older health maintenance organization members: A 2-year randomized trial of the cooperative health care clinic. *J Am Geriatrics Soc*. 2004; 52:1463–1470.
11. Freitag CM, Cholemkery H, Elsuni L, et al. The group-based social skills training SOSTA-FRA in children and adolescents with high functioning autism spectrum disorder – study protocol of the randomised, multi-centre controlled SOSTA – net trial. *Trials*. 2013; 14:12. [PubMed: 23302264]
12. Barlow J, Smailagic N, Huband N, et al. Group-based parent training programmes for improving parental psychosocial health. *Cochrane Database Systematic Rev*. 2012:200.
13. Roberts C, Roberts SA. Design and analysis of clinical trials with clustering effects due to treatment. *Clin Trials*. 2005; 2:152–162. [PubMed: 16279137]
14. Bauer DJ, Sterba SK, Hallfors DD. Evaluating group-based interventions when control participants are ungrouped. *Multivariate Behavior Res*. 2008; 43:210–236.
15. Baldwin SA, Bauer DJ, Stice E, et al. Evaluating models for partially clustered designs. *Psychol Meth*. 2011; 16:149–165.
16. Moerbeek M, Wong WK. Sample size formulae for trials comparing group and individual treatments in a multilevel model. *Stat Med*. 2008; 27:2850–2864. [PubMed: 17960589]
17. Cohen, J. Statistical power analysis for the behavioral science. Hillsdale, NJ: Lawrence Erlbaum Associates; 1988.
18. Kenward MG, Roger JH. Small sample inference for fixed effects from restricted maximum likelihood. *Biometrics*. 1997; 53:983–997. [PubMed: 9333350]
19. Diggle, PJ.; Heagerty, P.; Linag, K-Y., et al. Analysis of longitudinal data. 2. New York: Oxford University Press; 2002.
20. Heo M, Leon AC. Statistical power and sample size requirements for three level hierarchical cluster randomized trials. *Biometrics*. 2008; 64:1256–1262. [PubMed: 18266889]
21. Teerenstra S, Moerbeek M, van Achterberg T, et al. Sample size calculations for 3-level cluster randomized trials. *Clinical Trials*. 2008; 5:486–495. [PubMed: 18827041]
22. Davis K, Abrams M, Stremikis K. How the affordable care act will strengthen the nation’s primary care foundation. *J Gen Intern Med*. 2011; 26:1201–1203. [PubMed: 21523495]

Table 1

Comparison of theoretical power and simulation-based empirical power when group sizes are equal in a two-level model: Determinations of J with given K .

(2)	ρ	Experimental arm: $N_E = JK$			Control arm: $N_C = J'$			$\phi^{(2)}$	$\tilde{\phi}^{(2)}$
		J	K		$J' = JK$	$N_E + N_C$			
0.4	0.2	18	10		180	360		0.807	0.802
	0.4	26	10		260	520		0.807	0.805
	0.6	34	10		340	680		0.807	0.796
0.5	0.2	26	5		130	260		0.807	0.807
	0.4	32	5		160	320		0.807	0.813
	0.6	38	5		190	380		0.807	0.818
0.6	0.2	12	10		120	240		0.823	0.798
	0.4	17	10		170	340		0.816	0.793
	0.6	22	10		220	440		0.812	0.794
0.6	0.2	17	5		85	170		0.816	0.810
	0.4	21	5		105	210		0.817	0.796
	0.6	24	5		120	240		0.802	0.790
0.6	0.2	8	10		80	160		0.807	0.786
	0.4	12	10		120	240		0.822	0.792
	0.6	15	10		150	300		0.805	0.786
0.6	0.2	12	5		60	120		0.822	0.800
	0.4	14	5		70	140		0.801	0.780
	0.6	17	5		85	170		0.810	0.776
Mean								0.811	0.797

Note: (2) is a standardized effects size for two-level models; ρ (2) is the intra-class correlation coefficient of outcome Y within groups in the experimental arm; J is the number of groups in the experimental arm determined based on equation (6); K is the given number of subjects per group; $\phi^{(2)}$ is the theoretical power based on equation (5); and $\tilde{\phi}^{(2)}$ is the empirical power estimated based on 1000 simulations for each combination of design parameters.

Table 2

Comparison of theoretical power and simulation-based empirical power when group sizes are equal in a two-level model: Determinations of K with given J .

(2)	Experimental arm: $N_E = JK$			Control arm: $N_C = J'$		
	ρ	J	K	$J' = JK$	$N_E + N_C$	$\phi^{(2)}$
0.4	0.025	5	26	130	260	0.807
	0.050	5	37	185	370	0.802
	0.075	5	69	345	690	0.800
0.5	0.025	5	15	75	150	0.811
	0.050	5	18	90	180	0.809
	0.075	5	22	110	220	0.800
0.6	0.025	5	10	50	100	0.816
	0.050	5	11	55	110	0.811
	0.075	5	12	60	120	0.800
Mean						0.806

Note: (2) is a standardized effects size for two-level models; ρ (2) is the intra-class correlation coefficient of outcome Y within groups in the experimental arm; J is the given number of groups in the experimental arm; K is the number of subjects per group determined based on equation (7); $\phi^{(2)}$ is the theoretical power based on equation (5); and $\hat{\phi}^{(2)}$ is the empirical power estimated based on 1000 simulations for each combination of design parameters.

Table 3

Comparison of theoretical power and simulation-based empirical power when group sizes vary in a two-level model with varying sizes of K .

Experimental arm: $N_E = J \times \text{mean}(K_j)$				Control arm: $N_C = J'$		
ρ	J	K_j	$N_E + N_C$	J'	$\rho^{(2)}$	$\hat{\rho}^{(2)}$
0.4	18	U(5,15)	360	180	0.807	0.780
	26	U(5,15)	520	260	0.807	0.775
	34	U(5,15)	680	340	0.807	0.786
0.4	26	U(3,7)	260	130	0.807	0.808
	32	U(3,7)	320	160	0.807	0.782
	38	U(3,7)	380	190	0.807	0.775
0.5	12	U(5,15)	240	120	0.823	0.809
	17	U(5,15)	340	170	0.816	0.809
	22	U(5,15)	440	220	0.812	0.789
0.6	17	U(3,7)	170	85	0.816	0.825
	21	U(3,7)	210	105	0.817	0.798
	24	U(3,7)	240	120	0.802	0.779
0.6	8	U(5,15)	160	80	0.807	0.773
	12	U(5,15)	240	120	0.822	0.763
	15	U(5,15)	300	150	0.805	0.760
0.6	12	U(3,7)	120	60	0.822	0.803
	14	U(3,7)	140	70	0.801	0.767
	17	U(3,7)	170	85	0.810	0.752
Mean					0.811	0.785

Note: $U(a, b)$ denotes a uniform distribution with minimum a and maximum b ; $\rho^{(2)}$ is a standardized effects size for two-level models; $\hat{\rho}^{(2)}$ is the intra-class correlation coefficient of outcome Y within groups in the experimental arm; J is the number of groups in the experimental arm determined based on equation (6) for given $\text{mean}(K_j)$; $\rho^{(2)}$ is the theoretical power based on equation (5) with K replaced by $\text{mean}(K_j)$; and $\hat{\rho}^{(2)}$ is the empirical power estimated based on 1000 simulations for each combination of design parameters.

Table 4

Comparison of theoretical power and simulation-based empirical power when both center and group sizes are equal in a three-level model:
Determinations of K with given I and J .

(3)	Experimental arm: $N_E = IJK$					Control arm: $N_C = I'JK$					$N_E + N_C$	$\rho^{(3)}$	$\hat{\rho}^{(3)}$
	ρ_2	ρ_1	I	J	K	$I' = I$	JK						
0.4	0.1	0.4	14	5	10	14	50	1400	0.802	0.809			
		0.6	16	5	10	16	50	1600	0.812	0.830			
	0.2	0.4	23	5	10	23	50	2300	0.804	0.787			
		0.6	25	5	10	25	50	2500	0.811	0.810			
	0.1	0.4	13	10	5	13	50	1300	0.816	0.829			
		0.6	14	10	5	14	50	1400	0.827	0.819			
0.5	0.2	0.4	22	10	5	22	50	2200	0.804	0.807			
		0.6	23	10	5	23	50	2300	0.811	0.802			
	0.1	0.4	9	5	10	9	50	900	0.804	0.784			
		0.6	10	5	10	10	50	1000	0.803	0.776			
	0.2	0.4	15	5	10	15	50	1500	0.811	0.805			
		0.6	16	5	10	16	50	1600	0.811	0.807			
0.6	0.1	0.4	8	10	5	8	50	800	0.801	0.781			
		0.6	9	10	5	9	50	900	0.829	0.839			
	0.2	0.4	14	10	5	14	50	1400	0.802	0.795			
		0.6	15	10	5	15	50	1500	0.818	0.826			
	0.1	0.4	7	5	10	7	50	700	0.846	0.834			
		0.6	7	5	10	7	50	700	0.806	0.807			
	0.2	0.4	11	5	10	11	50	1100	0.832	0.824			
		0.6	11	5	10	11	50	1100	0.807	0.796			
	0.1	0.4	6	10	5	6	50	600	0.831	0.811			
		0.6	6	10	5	6	50	600	0.813	0.805			
	0.2	0.4	10	10	5	10	50	1000	0.813	0.795			
		0.6	10	10	5	10	50	1000	0.802	0.789			
Mean											0.813	0.807	

Note: (3) is a standardized effects size for three-level models; ρ_1 (11) is the correlation among the subject-level observations of outcome Y in the experimental arm; ρ_2 (10) is the correlation among the group-level observations of outcome Y in the experimental arm; I is the number of centers in the experimental arm determined based on equation (17); J is the given number of groups per center;

K is the given number of subjects per group; and $\phi^{(3)}$ is the theoretical power based on equation (16); $\phi^{(3)}$ is the empirical power estimated based on 1000 simulations for each combination of design parameters.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Table 5

Comparison of theoretical power and simulation-based empirical power when both center and group sizes are equal in a three-level model: Determinations of I with given J and K .

(3)	Experimental arm: $N_E = IJK$				Control arm: $N_C = I'JK$				$N_E + N_C$	$\phi^{(3)}$	$\tilde{\phi}^{(3)}$
	ρ_2	ρ_1	I	J	K	$I' = I$	JK				
0.4	0.01	0.1	5	5	6	5	30	300	0.815	0.787	
		0.2	5	5	8	5	40	400	0.815	0.784	
	0.02	0.1	5	5	8	5	40	400	0.804	0.800	
		0.2	5	5	13	5	65	650	0.805	0.779	
0.5	0.01	0.1	5	5	3	5	15	150	0.803	0.793	
		0.2	5	5	4	5	20	200	0.853	0.815	
	0.02	0.1	5	5	4	5	20	200	0.833	0.801	
		0.2	5	5	4	5	20	200	0.808	0.787	
0.6	0.01	0.1	5	5	2	5	10	100	0.820	0.777	
		0.2	5	5	2	5	10	100	0.820	0.770	
	0.02	0.1	5	5	3	5	15	150	0.892	0.855	
		0.2	5	5	3	5	15	150	0.881	0.864	
Mean											
0.813 0.807											

Note: (3) is a standardized effects size for three-level models; ρ_1 (11) is the correlation among the subject-level observations of outcome Y in the experimental arm; ρ_2 (10) is the correlation among the group-level observations of outcome Y in the experimental arm; I is the given number of centers in the experimental arm; J is the given number of groups per center; K is the number of subjects per group determined based on equation (19); $\phi^{(3)}$ is the theoretical power based on equation (16); and $\tilde{\phi}^{(3)}$ is the empirical power estimated based on 1000 simulations for each combination of design parameters.

Comparison of theoretical power and simulation-based empirical power when both center and group sizes vary in a three-level model: Determinations of I with varying sizes of J and K .

Table 6

(3)	Experimental arm: $N_E = I \times \text{mean}(J_i) \times \text{mean}(K_{ij})$					Control arm: $N_C = I' \times \text{mean}(K_i)$			$N_E + N_C$	$\phi^{(3)}$	
	ρ_2	ρ_1	I	J_i	K_{ij}	$I' = I$	K_i				
0.4	0.1	0.4	14	U(3,7)	U(5,15)	14	U(25,75)	1400	0.802	0.791	
		0.6	16	U(3,7)	U(5,15)	16	U(25,75)	1600	0.812	0.821	
	0.2	0.4	23	U(3,7)	U(5,15)	23	U(25,75)	2300	0.804	0.786	
		0.6	25	U(3,7)	U(5,15)	25	U(25,75)	2500	0.811	0.794	
	0.1	0.4	13	U(5,15)	U(3,7)	13	U(25,75)	1300	0.816	0.795	
		0.6	14	U(5,15)	U(3,7)	14	U(25,75)	1400	0.827	0.802	
0.5	0.2	0.4	22	U(5,15)	U(3,7)	22	U(25,75)	2200	0.804	0.803	
		0.6	23	U(5,15)	U(3,7)	23	U(25,75)	2300	0.811	0.794	
	0.1	0.4	9	U(3,7)	U(5,15)	9	U(25,75)	900	0.804	0.781	
		0.6	10	U(3,7)	U(5,15)	10	U(25,75)	1000	0.803	0.802	
	0.2	0.4	15	U(3,7)	U(5,15)	15	U(25,75)	1500	0.811	0.802	
		0.6	16	U(3,7)	U(5,15)	16	U(25,75)	1600	0.811	0.801	
0.6	0.1	0.4	8	U(5,15)	U(3,7)	8	U(25,75)	800	0.801	0.783	
		0.6	9	U(5,15)	U(3,7)	9	U(25,75)	900	0.829	0.811	
	0.2	0.4	14	U(5,15)	U(3,7)	14	U(25,75)	1400	0.802	0.790	
		0.6	15	U(5,15)	U(3,7)	15	U(25,75)	1500	0.818	0.812	
	0.1	0.4	7	U(3,7)	U(5,15)	7	U(25,75)	700	0.846	0.831	
		0.6	7	U(3,7)	U(5,15)	7	U(25,75)	700	0.806	0.814	
	0.2	0.4	11	U(3,7)	U(5,15)	11	U(25,75)	1100	0.832	0.808	
		0.6	11	U(3,7)	U(5,15)	11	U(25,75)	1100	0.807	0.783	
	0.1	0.4	6	U(5,15)	U(3,7)	6	U(25,75)	600	0.831	0.828	
		0.6	6	U(5,15)	U(3,7)	6	U(25,75)	600	0.813	0.787	
	0.2	0.4	10	U(5,15)	U(3,7)	10	U(25,75)	1000	0.813	0.783	
		0.6	10	U(5,15)	U(3,7)	10	U(25,75)	1000	0.802	0.786	
Mean										0.813	0.800

Note: $U(a, b)$ denotes a uniform distribution with minimum a and maximum b ; ρ_1 (11) is the correlation among the subject-level observations of outcome Y in the experimental arm; ρ_2 (10) is the correlation among the group-level observations of outcome Y in the experimental arm; I is the number of centers in the experimental arm determined based on equation (17) for the given mean of I_j and the given mean of K_{ij} ; $\phi^{(3)}$ is the theoretical power based on equation (16) with I and K replaced by $\text{mean}(I_j)$ and $\text{mean}(K_{ij})$, respectively; and $\hat{\phi}^{(3)}$ is the empirical power estimated based on 1000 simulations for each combination of design parameters.