Published in final edited form as:
Comput Stat Data Anal. 2013 April 1; 60: 169–178. doi:10.1016/j.csda.2012.11.016.

Sample size requirements to detect an intervention by time interaction in longitudinal cluster randomized clinical trials with random slopes

Moonseong Heo, **Xiaonan Xue**, and **Mimi Y. Kim**
Division of Biostatistics, Department of Epidemiology and Population Health, Albert Einstein College of Medicine, Bronx, New York, USA

Abstract

In longitudinal cluster randomized clinical trials (cluster-RCT), subjects are nested within a higher level unit such as clinics and are evaluated for outcome repeatedly over the study period. This study design results in a three level hierarchical data structure. When the primary goal is to test the hypothesis that an intervention has an effect on the rate of change in the outcome over time and the between-subject variation in slopes is substantial, the subject-specific slopes are often modeled as random coefficients in a mixed-effects linear model. In this paper, we propose approaches for determining the samples size for each level of a 3-level hierarchical trial design based on ordinary least squares (OLS) estimates for detecting a difference in mean slopes between two intervention groups when the slopes are modeled as random. Notably, the sample size is not a function of the variances of either the second or the third level random intercepts and depends on the number of second and third level data units only through their product. Simulation results indicate that the OLS-based power and sample sizes are virtually identical to the empirical maximum likelihood based estimates even with varying cluster sizes. Sample sizes for random versus fixed slope models are also compared. The effects of the variance of the random slope on the sample size determinations are shown to be enormous. Therefore, when between-subject variations in outcome trends are anticipated to be significant, sample size determinations based on a fixed slope model can result in a seriously underpowered study.

Keywords

longitudinal cluster RCT; three level data; power; sample size; random slope; effect size

1. Introduction

Longitudinal cluster randomized trials in which subjects are repeatedly assessed over time during the follow-up period as in a typical longitudinal cohort design (Feldman and McKinlay, 1994) can result in a three level data structure: the repeated measures (level 1) are nested within subjects (level 2) who in turn are nested within the randomized clusters, such as clinics (level 3). Often, the primary goal in these studies is to compare the

Address correspondence to: Moonseong Heo, Ph.D. Division of Biostatistics, Department of Epidemiology and Population Health, Albert Einstein College of Medicine, 1300 Morris Park Avenue, Belfer 13th Floor, Bronx, NY 10461, Phone (718) 920 6274, Fax (718) 515 6039, moonseong.heo@einstein.yu.edu.

longitudinal trends in a continuous outcome between the control and experimental groups. For example, in a cluster randomized trial to evaluate the effect of an intervention for depression in the primary care setting, clinics were randomly assigned to a control or intervention group and patients were repeatedly evaluated during follow up for change in depression symptoms using the Hamilton rating scale for depression (Alexopoulos et al., 2005; Dietrich et al., 2004). One of the study hypotheses was that the severity of depression symptoms would decline more rapidly among subjects who were treated at the primary care clinics assigned to the intervention compared to those who were treated with usual care.

Sample size and power calculations are essential in the proper design of a longitudinal cluster randomized trial. For a repeatedly measured outcome that is a continuous variable, data from the longitudinal cluster randomized trial can be analyzed by fitting a linear mixed effects model. The difference in mean slopes between intervention groups is assessed by including in the model an interaction term between the treatment and time effects. Heo and Leon (2009) proposed an approach for determining the required number of clusters to detect an intervention by time interaction under a fixed slope model. When the between-subject heterogeneity in the individual slopes is substantial, however, a random slope coefficients model may be more relevant (Laird and Ware, 1982; Longford, 1993; Murray et al., 2007).

Under this situation, Murray et al (2007) determined the minimum detectable effect size for testing mean slope differences in a more comprehensive situation allowing for cluster-specific random slopes. Roy et al (2007) also considered cluster- and subject-level random effects and presented implicit formulas for sample size requirements based on critical regions determined by χ^2 distributions of feasible version generalized least square estimates. Their sample determinations were based on approximate F distributions under alternative hypotheses. Preisser et al (2003) considered special cases in which only two time measurements, pre- and post-intervention, are considered. Their derivations were based on the pre-post difference using generalized estimating equation.

In this paper, we: 1) derive explicit closed form power functions and sample size formulae *for all levels* based on an ordinary least squares estimate (OLS) of the interaction effect under a subject-specific random slope model when subjects are measured multiple times during follow-up; 2) conduct an extensive simulation study to verify the statistical power achieved with the estimated sample sizes where the empirical statistical power is based on maximum likelihood estimates (MLE) considering varying cluster sizes and varying magnitudes of statistical power; and 3) compare sample sizes under the fixed and random slope coefficient models to assess the impact of the variance of the random slope on the sample size requirements. This allows one to evaluate the consequence in terms of power of designing a study using the fixed coefficient approach but fitting a random coefficient model in the actual analysis.

2. Statistical Model

A three level mixed-effects linear model for outcome Y with subject-specific random slopes can be expressed as follows (Hedeker and Gibbons, 2006):

$$Y_{ijk}=\beta_0+\xi X_{ijk}+\tau T_{ijk}+\delta X_{ijk}T_{ijk}+\nu_{j(i)}T_{ijk}+u_i+u_{j(i)}+e_{ijk},$$

where $i = 1, 2, \ldots, 2N_3$ is the index for the level three unit (e.g., clinic); $j = 1, \ldots, N_2$, is the index for the level two unit (e.g., subject) nested within each i; and $k = 1, 2, \ldots, N_1$, is the index for the level one unit (e.g., repeated outcomes) within each j. The intervention assignment indicator variable $X_{ijk} = 0$ and 1 if the i-th level three unit is assigned to a control intervention and an experimental intervention, respectively; therefore $X_{ijk} = X_i$ for

all j and k. Here we assume a balanced design so that $\Sigma_i X_i = N_3$. In addition, it is assumed that $T_{ijk} = T_k$ for all i and j, and that the time from $T_1 = 0$ (the baseline) to $T_{\text{end}} = N_1 - 1$ (the last time point) increases by equal unit time intervals. These assumptions reduce the model above to

$$Y_{ijk}=\beta_0+\xi X_i+\tau T_k+\delta X_i T_k+\nu_{j(i)} T_k+u_i+u_{j(i)}+e_{ijk} \quad (1)$$

With respect to the random effects, it is assumed that the error term e_{ijk} is normally distributed as $N\left(0,\sigma_e^2\right)$, the level two random intercept (i.e., subject-specific intercept) as $u_{j(i)}\sim N\left(0,\sigma_2^2\right)$, the level three random intercept (i.e., cluster-specific intercept) as $u_i\sim N\left(0,\sigma_3^2\right)$ and the random slope (i.e., subject-specific slope) as $\nu_{j(i)}\sim N\left(0,\sigma_\tau^2\right)$. Among these random components, it is further assumed that $u_i \perp u_{j(i)} \perp e_{ijk} \perp \nu_{j(i)}$, i.e., these four random components are mutually independent. In addition, *conditional independence* is assumed for all $u_{j(i)}$, $\nu_{j(i)}$ and e_{ijk}, whereas the u_i are *unconditionally* independent. That is, both $u_{j(i)}$ and $\nu_{j(i)}$ are independent over j conditional on u_i, and e_{ijk} are independent over k conditional on u_i, $\nu_{j(i)}$ and $u_{j(i)}$. This modeling framework was also considered in Murray et al (2007) which additionally allowed for cluster-specific random slopes. When the variance of the random slope is equal to 0 ($\sigma_\tau^2=0$), model (1) reduces to the fixed slope model which Heo and Leon (2009) previously considered. For the derivation of power function, we assume that all of these variances are known.

For the fixed effects, the parameter ξ represents the intervention effect at baseline, and the parameter τ represents the slope associated with the time effect, that is, the magnitude of the change in outcome over time, in the control group. Finally, the intervention-by-time effect δ, the parameter of primary interest, represents the difference in *mean* slopes of the outcome Y between the intervention groups. The overall intercept (fixed) is denoted by β_0.

Given that the parameter δ is of primary interest, the relevant null hypothesis can be expressed as:

$$H_0:\delta=0. \quad (2)$$

Under model (1), it can be shown that the elements of the mean vector for the outcome are equal to:

$$E(Y_{ijk})=\beta_0+\xi X_i+\tau T_k+\delta X_i T_k \quad (3)$$

and the elements of the covariance matrix are:

$$Cov\left(Y_{ijk}, Y_{i'j'k'}\right)=1(i=i'\cap j=j'\cap k=k')\sigma_e^2+1(i=i'\cap j=j')\left(T_k T_{k'}\sigma_\tau^2+\sigma_2^2\right), +1(i=i')\sigma_3^2 \quad (4)$$

where 1(.) is an indicator function (see Appendix A for a proof). It follows that:

$$Var\left(Y_{ijk}\right)=Cov\left(Y_{ijk}, Y_{ijk}\right)=\sigma^2+T_k^2\sigma_\tau^2, \quad (5)$$

where $\sigma^2 \equiv \sigma_e^2+\sigma_2^2+\sigma_3^2$, the variance of Y under the fixed slope model with $\sigma_\tau^2=0$. Therefore, the correlations among the level two data, i.e., among outcomes from different second level clusters (subjects) but the same third level cluster (clinic), can be expressed for $j \neq j'$ as follows:

$$Corr\left(Y_{ijk}, Y_{ij'k'}\right)=\frac{\sigma_3^2}{\sqrt{\sigma^2+T_k^2\sigma_\tau^2}\sqrt{\sigma^2+T_{k'}^2\sigma_\tau^2}}.$$

The correlations among the level one data, i.e., among outcomes measured at different time points on the same subject nested within clinics, can be expressed for $k \neq k'$ as:

$$Corr\left(Y_{ijk}, Y_{ijk'}\right)=\frac{\sigma_2^2+\sigma_3^2+T_kT_{k'}\sigma_\tau^2}{\sqrt{\sigma^2+T_k^2\sigma_\tau^2}\sqrt{\sigma^2+T_{k'}^2\sigma_\tau^2}}.$$

Under the fixed slope model, i.e., when $\sigma_\tau^2=0$, the correlations reduce to the following, respectively:

$$\rho_2=\sigma_3^2/\sigma^2 \quad (6)$$

and

$$\rho_1=(\sigma_2^2+\sigma_3^2)/\sigma^2. \quad (7)$$

3. Ordinary Least Square Estimate and its Variance

The ordinary least square (OLS) estimate $\hat{\delta}$ of the interaction effect is the difference in mean slopes between the two groups: that is,

$$\widehat{\delta}=\widehat{\eta}_1-\widehat{\eta}_0, \quad (8)$$

where $\hat{\eta}_g (g = 0,1)$ is the OLS estimate of the slope for the outcome Y in the g-th group in which $X_i = g$. Specifically, for i in the g-th group,

$$\begin{aligned}\widehat{\eta}_g&=\sum_{i=1}^{N_3}\sum_{j=1}^{N_2}\sum_{k=1}^{N_1}\left(T_k-\overline{T}\right)\left(Y_{ijk}-\overline{Y}_g\right)/\sum_{i=1}^{N_3}\sum_{j=1}^{N_2}\sum_{k=1}^{N_1}\left(T_k-\overline{T}\right)^2\\&=\sum_{i=1}^{N_3}\sum_{j=1}^{N_2}\sum_{k=1}^{N_1}\left(T_k-\overline{T}\right)\left(Y_{ijk}-\overline{Y}_g\right)/N_3N_2N_1Var_p(T),\end{aligned} \quad (9)$$

where: 1) $\overline{Y}_g (g = 0,1)$ is the overall group mean of the outcome Y for the g-th group; 2) $\overline{T}=\sum_{k=1}^{N_1}T_k/N_1$ is the "mean" time point; and 3) $Var_p(T)=\sum_{k=1}^{N_1}\left(T_k-\overline{T}\right)^2/N_1$ is the "population variance" the time variable T. Based on equations (4) and (5), it can be shown that the OLS estimate $\hat{\delta}$ is unbiased, i.e., $E(\hat{\delta}) = E(\hat{\eta}_1 - \hat{\eta}_0) = (\tau + \delta) - \tau = \delta$ (see Appendix B for a proof). Furthermore, the sampling distribution of OLS estimate $\hat{\delta}$ is normal since it is a linear combination of normally distributed Y_{ijk} and even if Y_{ijk}'s are correlated. In the present case with a perfectly balanced design, the OLS estimate (9) is the mean of subject-specific slope estimates whose large sample properties are similar to those of generalized least square estimates of the mean of random slopes even when the variance components are unknown (Gumpertz and Pantula, 1989).

The variance of $\hat{\eta}_g$ can be obtained based on equation (5) as follows (see Appendix C for a proof):

$$Var\left(\widehat{\eta}_g\right)=\frac{\sigma_e^2}{N_3N_2N_1Var_p(T)}+\frac{\sigma_\tau^2}{N_3N_2}. \quad (10)$$

It is notable that the variance of $\hat{\eta}_g$ does not depend on either the first or the second level random intercept, i.e., either σ_3^2 or σ_2^2. The variance of $\hat{\eta}_g$ can be expressed as:

$$Var\left(\widehat{\eta}_g\right)=\frac{(1-\rho_1)\sigma^2+N_1Var_p(T)\sigma_\tau^2}{N_3N_2N_1Var_p(T)}.$$

It follows that

$$Var\left(\widehat{\delta}\right)=Var\left(\widehat{\eta}_1-\widehat{\eta}_0\right)=Var\left(\widehat{\eta}_1\right)+Var\left(\widehat{\eta}_0\right)=\frac{2\left\{(1-\rho_1)\sigma^2+N_1Var_p(T)\sigma_\tau^2\right\}}{N_3N_2N_1Var_p(T)} \quad (11)$$

since $\hat{\eta}_1$ and $\hat{\eta}_0$ are independent of each other. Although the approaches to the derivations are different, this variance function reduces to that of equation (4) in Murray et al (2007) if the variance of the cluster-specific random slopes is zero in their model. This implies that the variance of the OLS estimate does not indeed depend on the variances of the two intercepts. Furthermore, equation (11) is a natural extension of equation (11) in Schlesselman et al. (1973) in which only two levels, subjects and repeated measures were considered.

4. Power and sample size

The following test statistic D, based on (8) and (11), can be used to test the null hypothesis (2):

$$D=\frac{\widehat{\delta}}{se(\widehat{\delta})}=\frac{(\widehat{\eta}_1-\widehat{\eta}_0)\sqrt{N_3N_2N_1Var_p(T)}}{\sqrt{2\left\{(1-\rho_1)\sigma^2+N_1Var_p(T)\sigma_\tau^2\right\}}}.$$

If the four variance components, σ_2^2, σ_3^2, σ_τ^2 and σ_e^2, are known, then the test statistic D is normally distributed with mean $\delta/se(\hat{\delta})$ and variance 1, since the estimated slope difference $\hat{\delta} = \hat{\eta}_1 - \hat{\eta}_0$ is normally distributed. Thus, under the null hypothesis (2), $D \sim N(0, 1)$ and under the alternative hypothesis of $\delta \neq 0$, $D \sim N(\delta/se(\hat{\delta}), 1)$.

The power of the test statistic D, denoted by ϕ, can therefore be written as follows:

$$\phi=1-\beta=\Phi\left\{\delta\sqrt{\frac{N_3N_2N_1Var_p(T)}{2\left\{(1-\rho_1)\sigma^2+N_1Var_p(T)\sigma_\tau^2\right\}}}-\Phi^{-1}(1-\alpha/2)\right\}, \quad (12)$$

where α is a two-sided significance level; β represents the probability of a type II error; Φ is the cumulative distribution function (CDF) of a standard normal distribution and Φ^{-1} is its inverse. We assume that: 1) $\delta = |\delta| > 0$; and 2) the probability below a critical value, $\Phi^{-1}(\alpha/2)$, in the other side under the alternative hypothesis is negligible and thus assumed to be 0.

The effect size is defined as:

$$\Delta=\delta/\sigma \quad (13)$$

and the ratio of the random slope variance to the sum of the other variances as

$$r_\tau=\sigma_\tau^2/\left(\sigma_2^2+\sigma_3^2+\sigma_e^2\right)=\sigma_\tau^2/\sigma^2. \quad (14)$$

Then, the power function (12) can be re-expressed as follows:

$$\phi=\Phi\left\{\Delta\sqrt{\frac{N_3N_2N_1Var_p(T)}{2\left\{(1-\rho_1)+r_\tau N_1 Var_p(T)\right\}}}-\Phi^{-1}(1-\alpha/2)\right\}. \quad (15)$$

When $\sigma_\tau^2=0$ or $r_\tau = 0$, the effect size Δ is identical to the standardized effect size for the slope difference δ and the power function (15) reduces to that derived by Heo and Leon under a fixed slope model. The power function increases with Δ (13) and ρ_1 (7) but decreases with r_τ (14) or with the random slope variance σ_τ^2.

It follows that when hypothesis testing is based on D with a two-sided significance level α, the required sample size per group for the third level unit N_3, for a desired statistical power ϕ can be calculated from equation (15) as:

$$N_3=\frac{2\left\{(1-\rho_1)+r_\tau N_1 Var_p(T)\right\}\left\{\Phi^{-1}(1-\alpha/2)+\Phi^{-1}(1-\beta)\right\}^2}{N_2N_1Var_p(T)\Delta^2}. \quad (16)$$

More precisely, N_3 is the smallest integer greater than the right hand side of equation (16). Note that the level 3 sample size is inversely associated with ρ_1 (7) and N_2, the sample size of the level 2 unit. It follows that N_2 can be expressed as:

$$N_2=\frac{2\left\{(1-\rho_1)+r_\tau N_1 Var_p(T)\right\}\left\{\Phi^{-1}(1-\alpha/2)+\Phi^{-1}(1-\beta)\right\}^2}{N_3N_1Var_p(T)\Delta^2}. \quad (17)$$

However, the sample size for the level one data, N_1, needs to be determined in an iterative manner because $Var_p(T)$ is a function of N_1. Specifically, an iterative solution for N_1 must satisfy the following equation:

$$N_1=\frac{(1-\rho_1)}{Var_p(T)\left[N_3N_2\Delta^2/\left\{2(\Phi^{-1}(1-\alpha/2)+\Phi^{-1}(1-\beta))^2\right\}-r_\tau\right]}. \quad (18)$$

5. Simulation study

We conducted simulation studies to verify the sample size N_3 (16) and the power function (15) under perfectly balanced design and under varying cluster sizes as well. We first determined N_3 for a two-sided significance level $\alpha = 0.05$ and a nominal power $\phi = 0.8$ under the following combinations: $\Delta T_{\text{end}} = \Delta(N_1 - 1) = 0.4, 0.5, 0.6$; $N_2 = 10, 20$; $N_1 = 5, 7, 9$; $\rho_1 = 0.4, 0.6$; $r_\tau = 0.1, 0.2$, while without loss of generality, $\sigma = 1$, $\rho_2 = 0.2$, $\beta_0 = \xi = 0$,

and $\tau = -1$ in model (1) remained fixed. Values for σ_3^2 and σ_2^2 were determined through ρ_2 (6) and ρ_1 (7). The full factorial number of combinations of the simulation parameters is 72. The effect size of the interaction Δ is specified as a standardized between-group mean difference $\Delta T_{end} = \Delta(N_1 - 1)$ at the end of trial under a fixed slope model. Effect sizes in the range of 0.4–0.6 have generally been referred to as medium (Cohen, 1988). We generated 1000 simulated data sets for each combination of the above parameters. Although the combinations of the parameters are somewhat arbitrarily selected, 1000 simulations are commonly considered in simulation studies (Burton et al., 2006).

Each simulated data set was produced by first estimating N_3 using equation (16) for $\phi = 0.8$ under a specific factorial combination of simulation parameters and then generating outcome data sets with the estimated N_3 in accordance to model (1): $Y_{ijk} = \beta_0 + \xi X_i + \tau T_k + \delta X_i T_k + \nu_{j(i)} T_k + u_i + u_{j(i)} + e_{ijk}$. We fit this model using SAS PROC MIXED with the maximum likelihood estimation option and retained the resulting p-values for testing the null hypothesis (2). However, all variance components are assumed to be unknown to reflect real data analysis situations. We denoted the p-value by $p_s(\delta)$ for the s-th simulated data set (s = 1, 2, .., 1000) and computed the empirical power ϕ as follows:

$$\phi = \sum_{s=1}^{1000} 1\{p_s(\delta) < \alpha)\} / 1000. \quad (19)$$

This empirical power was compared with the theoretical power ϕ on which the estimated N_3 was based. We note that ϕ is never less than the pre-specified power of 0.8 since N_3 is the smallest integer greater than the right hand side of equation (16).

In addition, the procedures above were repeated to examine the validity of the derived sample size formulas under varying nominal statistical power ranging from 0.6 to 0.9 with selected combinations of the simulation parameters above. We also conducted simulations with varying clusters sizes (i.e., varying number of subjects across clusters) randomly drawn from uniform distributions for effect sizes ΔT_{end} ranging from 0.4 to 0.9. In this case, we first determined N_2 for a given number of clusters N_3 (instead of determining N_3 for given N_2) and then considered a uniform random variable $U(a, b)$ with expectation N_2 to draw varying cluster sizes N_2; the integer values of a and b were determined as follows: a = N_2 − floor($3N_2$ /4) and b = N_2 + floor($3N_2$ /4) so that $a > 0$ and $E\{U(a, b)\} = (a + b)/2 = N_2$, where floor(x) returns the greatest integer smaller than or equal to x.

6. Simulation study results

Table 1 summarizes the empirical power ϕ (19) and the theoretical power ϕ (15) based on the estimated N_3 which was determined from equation (16) and the assumed simulation parameters above for a nominal 80% statistical power. The simulation-based empirical power estimates are virtually identical to the theoretical power as reflected by the negligible differences in the average values shown in the last row of Table 1. Furthermore, among the results from the 72 different combinations of the simulation parameters (Table 1), only two (<5%) of the absolute differences, $|\phi - \phi|$, were beyond the 95% confidence limit, $\pm 1.96\sqrt{0.8 \times 0.2/1000} = \pm 0.025$. Thus, the proposed formulae for sample size and power are very accurate under the conditions that were considered. In every case, the theoretical power is no less than 0.8, since the power calculations were based on “integer” value of N_3.

As one would expect, the required sample size for N_3 decreases with increasing effect size Δ for a given level of power. It also decreases with increasing correlation ρ_1; for example, when $N_2 = 5$, $N_1 = 10$, and $\Delta T_{end} = 0.4$, (or $\Delta = 0.4/4 = 0.1$) the respective sample sizes

requirements for 80% power for the level three data (N_3), were 26, and 22 for $\rho_1 = 0.4$ and 0.6. Furthermore, the theoretical power is identical for various combinations of N_2 and N_3 that yield the same product. For example, each of the following pairs of (N_2, N_3) with the same product of 100: ($N_2 = 10$, $N_3 = 10$) and ($N_2 = 20$, $N_3 = 5$) yielded identical power of 0.809 when $N_1 = 5$, $\rho_1 = 0.4$, $\Delta T_{end} = 0.6$ (or $\Delta = 0.6/4 = 0.15$) (Table 1). Therefore, the sample sizes are exchangeable between N_2 and N_3. However, the effect of r_τ is substantial in that N_3 increases by >~50% – <~100% when r_τ ranges from 0.1–0.2 under the conditions considered here.

Table 2 presents results under varying statistical power and shows that the empirical power is very close to theoretical power regardless of the nominal statistical power and other parameters. Results presented in Table 3 confirm that the derived sample size formulas are valid even when clusters sizes are unbalanced as long as as an average cluster size is equal to N_2 computed based on equations (17).

7. Comparison of sample sizes between random and fixed slope models

To systematically examine the effect of the slope variance ratio r_τ on N_3, we defined a sample size ratio $R(N_3)$ between random fixed slope models based on equation (16) as follows:

$$R(N_3; r_\tau, N_1, \rho_1) \equiv \frac{N_3|r_\tau>0}{N_3|r_\tau=0} = \frac{(1-\rho_1)+r_\tau N_1 Var_p(T)}{(1-\rho_1)} = 1+r_\tau N_1 Var_p(T)/(1-\rho_1).$$

Of note, the ratio is not a function of the effect size Δ. Under the assumption that the value of T increases from 0 to $T_{end} = N_1 - 1$ by unit time increments, the population variance of T reduces to $Var_p(T) = (N_1 - 1)(N_1 + 1)/12$ which yields:

$$R(N_3) = 1+r_\tau N_1 (N_1^2-1)/\{12(1-\rho_1)\}. \quad (20)$$

Thus, $R(N_3)$ is an increasing function of r_τ, ρ_1 and N_1. Table 4 shows that the effect of N_1 on the sample size ratio $R(N_3)$ is greatest for larger r_τ because the variance of the outcome increases quadratically with N_1 or T (4) and the magnitude of the increase in the variance is larger for larger r_τ.

8. Discussion

The power function (15) derived using the OLS estimates (9) was shown in our simulation studies to be accurate compared to the empirical power based on the maximum likelihood estimates (MLE) even with unknown variances and varying cluster sizes N_2, and varying nominal statistical power (Tables 1–3). Although we did not formally demonstrate the equivalence between the OLS and MLE approaches for estimating the intervention-by-time interaction effect δ the two methods yielded identical unbiased estimates in our simulation studies. Furthermore, another simulation study with the same combination of the parameters as in Table 1 using restricted maximum likelihood estimates (REML) yielded virtually identical statistical power (results not shown). In addition, differences in the standard errors obtained based on between equation (11) and the MLE approach were negligible. Furthermore, the standard error estimates were also unbiased (results not shown). All together, the size of the test statistic D is unbiased.

The derived explicit sample size formulas were, therefore, valid under various nominal statistical power (Table 2), and thus may be able to be readily applied for any combination

of model parameters beyond those considered in the simulation study (section 6). Therefore, the OLS based power function and sample size approach with known variance components can be applied to designing a longitudinal cluster-RCT that will be analyzed based on a mixed effects linear model with random slopes applying restricted or unrestricted maximum likelihood parameter estimation with unknown variance components. Furthermore, the sample size should be applicable to the case of a two-level data structure by replacing $N_3 = 1$ in equation (17) (Schlesselman, 1973).

It follows that the proposed approach is also accurate for determining the required sample sizes for any of the levels in the 3-level longitudinal cluster randomized trial as shown in Table 3. Through the ratio r_τ, the effect of the variance of the random slope on N_3 is enormous (Table 2) especially for larger N_3, i.e., for longer trials. For example, the ratio is as high as 27 even for small $r_\tau = 0.1$ and $\rho_1 = 0.3$ when $N_3 = 13$. This finding shows that designing a study using the fixed coefficient approach can substantially be underpowered if the between-subject variability in longitudinal trends are anticipated to be significant. Therefore, the inclusion and exclusion criteria for enrolling subjects into the trial should be carefully taken into consideration to minimize the heterogeneity in the subjects' anticipated outcome trajectories. It the between-subject variability in slopes is expected to be negligible based on the results of pilot studies or knowledge acquired through clinical experience, the fixed slope sample size formula in (Heo and Leon, 2009), which is a special case of equation (16) with $r_\tau = 0$, can be applied. However, when the time interval is not necessarily unity unlike the situation considered in this paper, increasing N_1 for a given duration does indeed reduce N_3; that is, more observations per subject for a given time frame increase the statistical power.

The effects of the two correlations ρ_1 and ρ_2, often referred to as intra class correlations (ICC), on the sample size determinations depend in general on the parameters which will be tested and the statistical models as well. Even though ρ_1 and ρ_2 have negative and no effect on the sample size, respectively, in the current situation, their effects on the sample size are not trivial for testing a main intervention effect (Heo and Leon, 2008; Teerenstra et al., 2008). Therefore, careful attention should be paid to the assessment of ICC in the context of the study objectives of the clinical trial (Campbell et al., 2005; Resnicow et al., 2010).

Although the intercept between the two groups should virtually be identical (i.e., $\xi = 0$) due to random allocation of the interventions, the sample size approach does not necessarily require $\xi = 0$ because the slope difference is independent from the intercept. Furthermore, the statistical model in general does not have to require that the two intercepts meet at baseline. Nevertheless, the condition $\xi = 0$ may be necessary, if not essential, for determination of the reference effect size.

The statistical power function depends on the number of second and third level data units only through their product, i.e., N_3N_2 (equations 16 and 17). Therefore, as far as testing the intervention by time interaction is concerned, the recruitment plan can be very flexible, which is useful in designing a longitudinal cluster RCT. For example, recruitment of $N_3N_2 = 100$ subjects per group would yield identical statistical power for the following different combinations of sample sizes: $N_2 = 25$ subjects per clinic and $N_3 = 4$ clinics; $N_2 = 10$ subjects and $N_3 = 10$ clinics; $N_2 = 5$ subjects and $N_3 = 20$ clinics. Therefore, one can design a cluster RCT in a variety of research settings given that different combinations of number of clinics and subjects can yield the same level of power. In this paper, however, it is implicitly assumed that $N_3 > 2$. When $N_3 = 2$, the sample size approach proposed by Preisser et al (2003) could also be applied although their approach does not necessarily coincide with ours because: 1) a different correlation structure for the outcome variables was assumed; and 2) the pre-post difference was not scaled by the corresponding time difference.

Model (1) requires the fewest number of parameters in a class of random slope models for analysis of three level data from a cluster-longitudinal study. However, it does not necessarily reflect more complex situations such as when mutual and conditional independence should be relaxed. To this end, other factors such as alternative within-subject correlation structures (e.g., autocorrelation structure), attrition problems (Roy et al., 2007) and costs associated with recruitment (Konstantopoulos, 2009) should clearly be taken into consideration for an optimal design of a cluster-RCT (Raudenbush and Liu, 2000). However, these important issues are beyond the scope of the present paper. To apply our sample size under an autoregressive correlation structure which assumes that the correlation decreases with time, the average correlation could be used in lieu of ρ_1. Nevertheless, it is unknown if this intuitive strategy would yield reasonably well approximated sample sizes. Roy et al (2007) extensively discussed the impact of attrition on the sample size with approximate F distributions for diverse situations: subject- and cluster-specific random non-linear time trends under both within and between cluster randomizations. In other multi-level trial situations the publically available software titled “Optimal Design Software”, available from the William T. Grant Foundation developed by Stephen Raudenbush and colleagues can be applied to determine sample sizes that allow for covariate adjustments which are based on noncentral F-distributions.

In conclusion, the proposed explicit and easily implementable formulae for sample sizes (16,17,18) and the power function (15) can be applied to designing cluster-randomized clinical trials even with varying cluster sizes that intend to compare *mean* slopes of outcomes ver time between two intervention groups in a three level data structure when between-subject variability in the slopes should be taken into consideration.

Acknowledgments

We are grateful to Dr. Xianhong Xie for his assistance with the simulation programming. We thank anonymous referees for their helpful comments that improve the contents of this manuscript. The present study was supported in part by the Albert Einstein Center for AIDS research grant P30 AI51519 and the Einstein-Montefiore Clinical and Translational Science Award UL1 RR025750.

References

Alexopoulos GS, Katz IR, Bruce ML, Heo M, Ten Have T, Raue P, Bogner HR, Schulberg HC, Mulsant BH, Reynolds CF. Remission in depressed geriatric primary care patients: A report from the PROSPECT study. A J Psychiatry. 2005; 162:718–724.

Burton A, Altman DG, Royston P, Holder RL. The design of simulation studies in medical statistics. Stat Med. 2006; 25:4279–4292. [PubMed: 16947139]

Campbell MK, Fayers PM, Grimshaw JM. Determinants of the intracluster correlation coefficient in cluster randomized trials: the case of implementation research. Clin Trials. 2005; 2:99–107. [PubMed: 16279131]

Cohen, J. Statistical Power Analysis for the Behavioral Science. Lawrence Erlbaum Associates; Hillsdale, NJ: 1988.

Dietrich AJ, Oxman TE, Williams JW, Schulberg HC, Bruce ML, Lee PW, Barry S, Raue PJ, Lefever JJ, Heo M, Rost K, Kroenke K, Gerrity M, Nutting PA. Re-engineering systems for the treatment of depression in primary care: cluster randomised controlled trial. Br Med J. 2004; 329:602–605. [PubMed: 15345600]

Feldman HA, McKinlay SM. Cohort versus cross-sectional design in large field trials - precision, sample-size, and a unifying model. Stat Med. 1994; 13:61–78. [PubMed: 9061841]

Gumpertz M, Pantula SG. A simple approach to inference in random coefficient models. Am Stat. 1989; 43:203–210.

Hedeker, D.; Gibbons, RD. Longitudinal Data Analysis. Wiley; Hoboken, NJ: 2006.

Heo M, Leon AC. Statistical Power and Sample Size Requirements for Three Level Hierarchical Cluster Randomized Trials. Biometrics. 2008; 64:1256–1262. [PubMed: 18266889]

Heo M, Leon AC. Sample size requirements to detect an intervention by time interaction in longitudinal cluster randomized clinical trials. Stat Med. 2009; 28:1017–1027. [PubMed: 19153969]

Konstantopoulos S. Incorporating Cost in Power Analysis for Three-Level Cluster-Randomized Designs. Evaluation Review. 2009; 33:335–357. [PubMed: 19509118]

Laird NM, Ware JH. Random-effects models for longitudinal data. Biometrics. 1982; 38:963–974. [PubMed: 7168798]

Longford, NT. Random Coefficient Models. Oxford University Press; New York: 1993.

Murray DM, Blitstein JL, Hannan PJ, Baker WL, Lytle LA. Sizing a trial to alter the trajectory of health behaviours: Methods, parameter estimates, and their application. Stat Med. 2007; 26:2297–2316. [PubMed: 17044139]

Preisser JS, Young ML, Zaccaro DJ, Wolfson M. An integrated population-averaged approach to the design, analysis and sample size determination of cluster-unit trials. Stat Med. 2003; 22:1235–1254. [PubMed: 12687653]

Raudenbush SW, Liu XF. Statistical power and optimal design for multisite randomized trials. Psychol Methods. 2000; 5:199–213. [PubMed: 10937329]

Resnicow K, Zhang NH, Vaughan RD, Reddy SP, James S, Murray DM. When Intraclass Correlation Coefficients Go Awry: A Case Study From a School-Based Smoking Prevention Study in South Africa. Am J Public Health. 2010; 100:1714–1718. [PubMed: 20167897]

Roy A, Bhaumik DK, Aryal S, Gibbons RD. Sample size determination for hierarchical longitudinal designs with differential attrition rates. Biometrics. 2007; 63:699–707. [PubMed: 17825003]

Schlesselman JJ. Planning a longitudinal study: II. Frequency of measurement and study duratio. J Chronic Dis. 1973; 26:561–570. [PubMed: 4759581]

Teerenstra S, Moerbeek M, van Achterberg T, Pelzer BJ, Borm GF. Sample size calculations for 3-level cluster randomized trials. Clin Trials. 2008; 5:486–495. [PubMed: 18827041]

Appendix A: Derivations of covariance and variance in equations (4) and (5)

First, we have $Cov(Y_{ijk}, Y_{i'j'k'}) = Cov(u_i + u_{j(i)} + \nu_{j(i)}T_k + e_{ijk}, u_{i'} + u_{j'(i')} + \nu_{j'(i')}T_{k'} + e_{i'j'k'}) = Cov(u_i, u_i) + Cov(u_{j(i)}, u_{j'(i')}) + Cov(\nu_{j(i)}T_k, \nu_{j'(i')}T_{k'}) + Cov(e_{ijk}, e_{i'j'k'})$. It is because covariances between the other pairs of terms are all 0 due to the mutual independence assumption. Second, $Cov(u_i, u_{i'}) = 1(i=i')\sigma_3^2$ since u_i are independent over i. Third, $Cov(u_{j(i)}, u_{j'(i')}) = 1(i=i' \cap j=j')\sigma_2^2$ since $u_{j(i)}$ and $u_{j'(i')}$ are independent: 1) regardless of equality of j and j' when $i \neq i'$; or 2) if $j \neq j'$ when $i = i'$ owing to the conditional independence. This reasoning with respect to application of the conditional independence assumption can be extended to the other covariances as follows: $Cov(\nu_{j(i)}T_k, \nu_{j'(i')}T_{k'}) = 1(i=i' \cap j=j')T_kT_{k'}\sigma_\tau^2$ and $Cov(e_{ijk}, e_{i'j'k'}) = 1(i=i' \cap j=j' \cap k=k')\sigma_e^2$. Therefore, equation (4) holds. Finally, $Cov(Y_{ijk}, Y_{i'j'k'}) = Var(Y_{ijk})$ by definition only when $i = i'$ and $j = j'$ and $k = k'$. Therefore, equation (5) holds.

Appendix B: Proof of unbiasedness of OLS estimate E(δ̂) = E(η̂1 − η̂0) = (τ + δ) − τ = δ

If $g = 0$, then $X_i = 0$ for all cluster i in group 0. Therefore, based on equation (3), we have $E(Y_{ijk}) = \beta_0 + \tau T_k$, $E(\bar{Y}_0) = \beta_0 + \tau\bar{T}$ and subsequently $E(Y_{ijk} - \bar{Y}_0) = \tau(T_k - \bar{T})$. It follows that $E(\hat{\eta}_0) = \tau$, i.e., we have based on equation (9):

$$E(\widehat{\eta}_0)=\sum_{i=1}^{N_3}\sum_{j=1}^{N_2}\sum_{k=1}^{N_1}\left(T_k-\overline{T}\right)E\left(Y_{ijk}-\overline{Y}_0\right)/\sum_{i=1}^{N_3}\sum_{j=1}^{N_2}\sum_{k=1}^{N_1}\left(T_k-\overline{T}\right)^2$$
$$=\tau\sum_{i=1}^{N_3}\sum_{j=1}^{N_2}\sum_{k=1}^{N_1}\left(T_k-\overline{T}\right)\left(T_k-\overline{T}\right)/\sum_{i=1}^{N_3}\sum_{j=1}^{N_2}\sum_{k=1}^{N_1}\left(T_k-\overline{T}\right)^2=\tau.$$

Similarly, if $g = 1$, then $X_i = 1$ for all cluster i in group 1. Therefore, $E(Y_{ijk}) = \beta_0 + \xi + \tau T_k + \delta T_k$, $E(\overline{Y}_1) = \beta_0 + \xi + \tau\bar{T} + \delta\bar{T}$ and subsequently $E(Y_{ijk} - \overline{Y}_1) = \tau(T_k - \bar{T}) + \delta(T_k - \bar{T})$. If follows again based on equation (9) that $E(\hat{\eta}_1) = \tau + \delta$. Therefore, $E(\hat{\delta}) = E(\hat{\eta}_1 - \hat{\eta}_0) = (\tau + \delta) - \tau = \delta$.

Appendix C: Proof of equation (10) of the sampling variance of η̂g, that is, Var(η^g)=σe2N3N2N1Varp(T)+στ2N3N2

Let $W_k = (T_k - \bar{T})$, then we have: $\sum_{k=1}^{N_1} W_k^2 = N_1 Var_p(T); \sum_{k=1}^{N_1} W_k = 0; \sum_{k'\neq k}^{N_1} W_{k'} = -W_k$; and

$$\widehat{\eta}_g=\sum_{i=1}^{N_3}\sum_{j=1}^{N_2}\sum_{k=1}^{N_1}W_k\left(Y_{ijk}-\overline{Y}_g\right)/N_3N_2N_1Var_p(T)=\sum_{i=1}^{N_3}\sum_{j=1}^{N_2}\sum_{k=1}^{N_1}W_kY_{ijk}/N_3N_2N_1Var_p(T).$$

Observing that Y is independent over i, we decompose the variance of the numerator of $\hat{\eta}_g$ as follows:

$$Var\left(\sum_{i=1}^{N_3}\sum_{j=1}^{N_2}\sum_{k=1}^{N_1}W_kY_{ijk}\right)=\underbrace{\sum_{i=1}^{N_3}\sum_{j=1}^{N_2}\sum_{k=1}^{N_1}W_k^2Cov\left(Y_{ijk},Y_{ijk}\right)}_{A}+\underbrace{\sum_{i=1}^{N_3}\sum_{j=1}^{N_2}\sum_{k=1}^{N_1}\sum_{k'\neq k}^{N_1}W_kW_{k'}Cov\left(Y_{ijk},Y_{ijk'}\right)}_{B}$$
$$+\underbrace{\sum_{i=1}^{N_3}\sum_{j=1}^{N_2}\sum_{j'\neq j}^{N_2}\sum_{k=1}^{N_1}\sum_{k'=1}^{N_1}W_kW_{k'}Cov\left(Y_{ijk},Y_{ij'k'}\right)}_{C}.$$

Now, recall equation (4), that is,

$$Cov\left(Y_{ijk},Y_{i'j'k'}\right)=1(i=i'\cap j=j'\cap k=k')\sigma_e^2+1(i=i'\cap j=j')\left(T_kT_{k'}\sigma_\tau^2+\sigma_2^2\right)+1(i=i')\sigma_3^2.$$

It follows that

$$A=\sigma^2N_3N_2N_1Var_p(T)+\sigma_\tau^2N_3N_2\sum\nolimits_k T_k^2(T_k-\overline{T})^2$$

since $Var(Y_{ijk})=\sigma^2+T_k^2\sigma_\tau^2$, where $\sigma^2 \equiv \sigma_e^2+\sigma_2^2+\sigma_3^2$. As for B, we have

$$B=\sigma_\tau^2N_3N_2\sum\nolimits_{k'\neq k}\sum\nolimits_k T_kT_{k'}(T_k-\overline{T})(T_{k'}-\overline{T})-(\sigma_2^2+\sigma_2^2)N_3N_2N_1Var_p(T).$$

The last term is due to $\sum_{k=1}^{N_1}\sum_{k'\neq k}^{N_1} W_k W_{k'} = -\sum_{k=1}^{N_1} W_k^2$ since $\sum_{k'\neq k}^{N_1} W_{k'} = -W_k$. It is easy to see that $C = 0$ since $\sum_{k=1}^{N_1} W_k = 0$. Hence, we have

$$Var\left(\sum_{i=1}^{N_3}\sum_{j=1}^{N_2}\sum_{k=1}^{N_1} W_k Y_{ijk}\right) = A + B$$
$$= (\sigma^2 - \sigma_2^2 - \sigma_3^2) N_3 N_2 N_1 Var_p(T) + \sigma_\tau^2 N_3 N_2 \left\{\sum_k T_k^2 (T_k - \overline{T})^2 + \sum_{k'\neq k}\sum_k T_k T_{k'} (T_k - \overline{T})(T_{k'} - \overline{T})\right\}$$
$$= \sigma_e^2 N_3 N_2 N_1 Var_p(T) + \sigma_\tau^2 N_3 N_2 \sum_{k'}\sum_k T_k T_{k'} (T_k - \overline{T})(T_{k'} - \overline{T})$$
$$= \sigma_e^2 N_3 N_2 N_1 Var_p(T) + \sigma_\tau^2 N_3 N_2 N_1^2 Var_p^2(T).$$

The last equation is due to:

$$\sum_{k'}\sum_k T_k T_{k'} (T_k - \overline{T})(T_{k'} - \overline{T}) = \left\{\sum_k T_k (T_k - \overline{T})^2\right\} = \left\{\sum_k T_k^2 - N_1 \overline{T}\right\}^2 = N_1^2 Var_p^2(T).$$

It follows that equation (10) above holds.

Table 1

Sample size N_3, theoretical power ϕ and empirical power ϕ for testing intervention group by time interaction effect in a three level mixed-effect *random slope coefficient* linear regression analysis, based on 1000 simulations: nominal statistical power = 0.8.

				$\Delta T_{end} = 0.4$			$\Delta T_{end} = 0.5$			$\Delta T_{end} = 0.6$		
r_τ	N_2	N_1	ρ_1	N_3	ϕ	ϕ	N_3	ϕ	ϕ	N_3	ϕ	ϕ
0.1	10	5	0.4	26	0.813	0.799	17	0.822	0.833	12	0.828	0.821
			0.6	22	0.800	0.817	15	0.825	0.818	10	0.809	0.797
		7	0.4	43	0.801	0.800	28	0.808	0.830	20	0.819	0.826
			0.6	41	0.806	0.804	26	0.803	0.806	18	0.801	0.789
		9	0.4	70	0.805	0.821	45	0.807	0.797	31	0.804	0.813
			0.6	67	0.800	0.791	43	0.801	0.790	30	0.803	0.812
	20	5	0.4	13	0.813	0.811	9	0.842	0.857	6	0.828	0.813
			0.6	11	0.800	0.803	8	0.848	0.849	5	0.809	0.818
		7	0.4	22	0.810	0.812	14	0.808	0.783	10	0.819	0.818
			0.6	21	0.815	0.819	13	0.803	0.786	9	0.801	0.801
		9	0.4	35	0.805	0.812	23	0.815	0.813	16	0.816	0.816
			0.6	34	0.806	0.807	22	0.810	0.803	15	0.803	0.792
0.2	10	5	0.4	41	0.802	0.823	27	0.813	0.803	19	0.818	0.811
			0.6	38	0.803	0.804	25	0.814	0.829	17	0.806	0.819
		7	0.4	79	0.804	0.804	51	0.807	0.809	35	0.803	0.794
			0.6	76	0.802	0.808	49	0.805	0.808	34	0.804	0.808
		9	0.4	132	0.800	0.800	85	0.803	0.821	59	0.803	0.790
			0.6	130	0.801	0.802	84	0.804	0.797	58	0.802	0.802
	20	5	0.4	21	0.811	0.831	14	0.827	0.835	10	0.837	0.838
			0.6	19	0.803	0.808	13	0.829	0.858	9	0.828	0.829
		7	0.4	40	0.809	0.806	26	0.815	0.805	18	0.814	0.810
			0.6	38	0.802	0.797	25	0.812	0.806	17	0.804	0.815
		9	0.4	66	0.800	0.801	43	0.807	0.844	30	0.809	0.813
			0.6	65	0.801	0.795	42	0.804	0.806	29	0.802	0.802
Mean					0.805	0.807		0.814	0.816		0.811	0.810

Comput Stat Data Anal. Author manuscript; available in PMC 2014 April 01.

N_1 = the number of level one units (repeated measures) per subjects; N_2 = the number of level two units (subjects) per clinic; N_3 = the number of level three units (clinics) per group, i.e., the sample size obtained from equation (16); $T_{end} = N_1 - 1$; ρ_1 = correlation among level one data under a fixed slope model (7); ρ = theoretical power based on the formula (15); ρ = empirical power based on equation (19); ρ = standardized effect size of the slope difference that yields an intervention effect ΔT_{end} at the end of a study; r_{τ} = the ratio of the random slope variance to the sum of the other variances (14), i.e., $r_\tau = \sigma_\tau^2 / \left(\sigma_2^2 + \sigma_3^2 + \sigma_e^2\right) = \sigma_\tau^2 / \sigma^2$.

Table 2

Sample size N_3, theoretical power ϕ and empirical power ϕ for testing with varying nominal statistical power intervention group by time interaction effect in a three level mixed-effect *random slope coefficient* linear regression analysis, based on 1000 simulations.

				Nominal statistical power								
				0.6			0.7			0.9		
ΔT_{end}	N_1	N_2	ρ_1	N_3	ϕ	ϕ	N_3	ϕ	ϕ	N_3	ϕ	ϕ
						$r_\tau = 0.1$						
0.4	5	10	0.4	16	0.609	0.567	20	0.705	0.697	34	0.903	0.906
		20	0.6	7	0.609	0.605	9	0.718	0.722	15	0.905	0.910
0.5	7	10	0.4	18	0.621	0.619	22	0.708	0.715	37	0.902	0.911
		20	0.6	9	0.648	0.637	11	0.734	0.729	18	0.911	0.909
0.6	9	10	0.4	20	0.618	0.598	25	0.715	0.720	42	0.906	0.912
		20	0.6	10	0.632	0.642	12	0.711	0.721	20	0.901	0.905
						$r_\tau = 0.2$						
0.4	5	10	0.4	26	0.609	0.622	33	0.712	0.719	55	0.902	0.894
		20	0.6	12	0.609	0.591	15	0.705	0.703	26	0.908	0.905
0.5	7	10	0.4	32	0.610	0.590	40	0.707	0.690	68	0.904	0.901
		20	0.6	16	0.624	0.641	20	0.721	0.721	33	0.905	0.896
0.6	9	10	0.4	37	0.605	0.601	47	0.708	0.716	79	0.902	0.906
		20	0.6	18	0.600	0.600	23	0.706	0.702	39	0.903	0.899
Mean					0.616	0.609		0.713	0.713		0.904	0.905

N_1 = the number of level one units (repeated measures) per subjects; N_2 = the number of level two units (subjects) per clinic; N_3 = the number of level three units (clinics) per group, i.e., the sample size obtained from equation (16); $T_{end} = N_1 - 1$; ρ_1 = correlation among level one data under a fixed slope model (7); ϕ = theoretical power based on the formula (15); ϕ = empirical power based on equation (19); Δ = standardized effect size of the slope difference that yields an intervention effect ΔT_{end} at the end of a study; r_τ = the ratio of the random slope variance to the sum of the other variances (14), i.e., $r_\tau = \sigma_\tau^2 / \left(\sigma_2^2 + \sigma_3^2 + \sigma_e^2\right) = \sigma_\tau^2 / \sigma^2$.

Table 3

Sample size N_2, theoretical power ϕ and empirical power ϕ for testing with varying cluster sizes drawn from uniform distributions and varying nominal statistical power intervention group by time interaction effect in a three level mixed-effect *random slope coefficient* linear regression analysis, based on 1000 simulations.

				Nominal statistical power											
				0.7				0.8				0.9			
ΔT_{end}	N_1	N_3	ρ_1	N_2	$U(a, b)$	ϕ	ϕ	N_2	$U(a, b)$	ϕ	ϕ	N_2	$U(a, b)$	ϕ	ϕ
							$r_\tau = 0.1$								
0.4	5	10	0.4	20	(5,35)	0.705	0.719	26	(7,45)	0.813	0.801	34	(9,59)	0.903	0.915
		20	0.6	9	(3,15)	0.718	0.713	11	(3,19)	0.800	0.783	15	(4,26)	0.905	0.906
0.5	7	10	0.4	22	(6,38)	0.708	0.696	28	(7,49)	0.808	0.788	37	(19,64)	0.902	0.906
		20	0.6	11	(3,19)	0.734	0.748	13	(4,22)	0.803	0.790	18	(5,31)	0.911	0.902
0.6	9	10	0.4	25	(7,43)	0.715	0.708	31	(8,54)	0.804	0.787	42	(11,73)	0.906	0.886
		20	0.6	12	(3,21)	0.711	0.699	15	(4,26)	0.803	0.800	20	(5,35)	0.901	0.900
							$r_\tau = 0.2$								
0.7	5	10	0.4	11	(3,19)	0.721	0.708	14	(4,24)	0.819	0.821	18	(5,31)	0.903	0.902
		20	0.6	5	(2,8)	0.714	0.714	7	(2,12)	0.848	0.844	9	(3,15)	0.923	0.923
0.8	7	10	0.4	16	(4,28)	0.717	0.726	20	(5,35)	0.809	0.807	27	(7,47)	0.909	0.896
		20	0.6	8	(2,14)	0.731	0.722	10	(3,17)	0.821	0.821	13	(4,22)	0.907	0.902
0.9	9	10	0.4	21	(6,36)	0.711	0.727	27	(7,47)	0.814	0.809	35	(9,61)	0.901	0.893
		20	0.6	11	(3,19)	0.737	0.735	13	(4,22)	0.806	0.815	18	(5,31)	0.913	0.903
Mean						0.719	0.718			0.812	0.806			0.907	0.903

N_1 = the number of level one units (repeated measures) per subjects; N_3 = the number of level three units (clinics) per group; N_2 = the number of level two units (subjects) per clinic; i.e., the sample size obtained from equation (17); $U(a, b)$ = uniform distribution with expectation N_2; $T_{end} = N_1 - 1$; ρ_1 = correlation among level one data under a fixed slope model (7); ϕ = theoretical power based on the formula (15); ϕ = empirical power based on equation (19); Δ = standardized effect size of the slope difference that yields an intervention effect ΔT_{end} at the end of a study; r_τ = the ratio of the random slope variance to the sum of the other variances (14), i.e., $r_\tau = \sigma_\tau^2 / \left(\sigma_2^2 + \sigma_3^2 + \sigma_e^2\right) = \sigma_\tau^2 / \sigma^2$.

Comput Stat Data Anal. Author manuscript; available in PMC 2014 April 01.

Table 4

Ratio of sample size N_3 between random and fixed slope: $R(N_3)$

		ρ_1		
r_τ	N_1	**0.3**	**0.5**	**0.7**
0.1	5	2.4	3.0	4.3
	9	9.6	13.0	21.0
	13	27.0	37.4	61.7
0.2	5	3.9	5.0	7.7
	9	18.1	25.0	41.0
	13	53.0	73.8	122.3
0.3	5	5.3	7.0	11.0
	9	26.7	37.0	61.0
	13	79.0	110.2	183.0

N_1 = the number of level one units (repeated measures) per subjects; ρ_1 = correlation among level one data under a fixed slope model (7); r_τ = the ratio of the random slope variance to the sum of the other variances (14), i.e., $r_\tau = \sigma_\tau^2 / \left(\sigma_2^2 + \sigma_3^2 + \sigma_e^2\right) = \sigma_\tau^2 / \sigma^2$.