Published in final edited form as:

Clin Trials. 2014 August ; 11(4): 503–507. doi:10.1177/1740774514532724.

Sample size requirements to detect a two- or three-way interaction in longitudinal cluster randomized clinical trials with second level randomization

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

Background—When randomizations are assigned at the cluster level for longitudinal cluster randomized trials (longitudinal-CRT) with a continuous outcome, formulae for determining the required sample size to detect a two-way interaction effect between time and intervention are available.

Purpose—To show that: 1) those same formulae can also be applied to longitudinal trials when randomizations are assigned at the subject level within clusters; and 2) this property can be extended to 2-by-2 factorial longitudinal-CRTs with two treatments and different levels of randomization for which testing a three-way interaction between time and the two interventions is of primary interest.

Methods—We show that slope estimates from different treatment arms are uncorrelated regardless of whether randomization occurs at the third or second level and also regardless of whether slopes are considered fixed or random in the mixed-effects model for testing two-way or three-way interactions. Sample size formulae are extended to unbalanced designs. Simulation studies were applied to verify the findings.

Results—Sample size formulae for testing two-way and three-way interactions in longitudinal-CRTs with second level randomization are identical to those for trials with third level randomization. In addition, the total number of observations required for testing a three-way interaction is demonstrated to be four times as large as that required for testing a two-way interaction regardless of level of randomization for both fixed and random slope models.

Limitations—The findings may be only applicable to longitudinal-CRTs with normally-distributed continuous outcome.

Conclusions—All of the findings are validated by simulation studies and enable the design of longitudinal clinical trials to be more flexible in regard to level of randomization and allocation of clusters and subjects.

Address correspondence to: Moonseong Heo, PhD, 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.

Conflict of Interest: None to declare.

Keywords

longitudinal CRT; three level data; power; sample size; within-cluster randomization

Introduction

In longitudinal cluster randomized trials (longitudinal-CRT), repeated measures (first level unit) are obtained during follow-up on study subjects (second level unit) who are nested within clusters (third level unit, e.g., clinics or hospitals), yielding a three level hierarchical data structure [1]. Under this type of design, a primary goal is to test whether the outcome trends over the follow-up period will be different between the experimental and control arms, i.e., to test the significance of a *two-way* interaction effect between time T and intervention X ($=0$ for control and $=1$ for experimental). Approaches for determining the required sample size to detect the XT interaction effect have been published for both fixed slope [2] and random slope models [3].

These papers showed that the power depends on N_2 and N_3 through N_2N_3 under fixed or random slope assumptions [2,3], where N_2 is the number of subject per cluster and N_3 is the number of clusters per arm. We call this the “invariance over product” property. Therefore, sample size determinations for the second and third units can be flexible for testing the XT interaction. For example, when $N_3N_2=100$ subjects per arm are needed for 80% power, sample sizes for N_3 and N_2 can be determined based on the availability of clusters and subjects. Recruitment of 10 subjects per clinic and 10 clinics per arm will result in the same power as recruitment of 20 subjects per clinic and 5 clinics per arm.

An important assumption of these sample size approaches is that the randomization occurs at the third level so that all the subjects within the cluster receive the same intervention, experimental or control. But in many settings, the randomization needs to be made at the subject level (i.e., second level) within clusters instead of at the third level across clusters. We note that the former type of design would be equivalent to a multi-center trial [4,5] with individual level randomization. In this case, the slope estimates might not necessarily be independent across treatment arms because data from subjects randomized to different arms within the same cluster are correlated, and this correlation may affect sample size calculations. For instance, sample size requirements for a main effect are quite different across randomizations at different levels [6]. Therefore, we formally examined the impact of randomization at the second (subject) level on the required sample size for testing the XT interaction.

Results

Even if randomization occurs at the subject as opposed to cluster level, the slope estimates within clusters are uncorrelated between arms and thus the variance of the slope differences is not affected by second level randomization for either the fixed or random slope model (see appendix for proof). It follows that the power functions and sample size formulae for longitudinal-CRTs with third level randomization still apply to longitudinal trials with second level randomization. We call this property “invariance over level of randomization.”

Specifically, a sample size formula for detecting a two-way interaction XT derived under a perfectly balanced design [2,3] with *randomization at the third level* can be extended to trials with $1:\lambda$ *unbalanced* allocations ($\lambda=1$ for balanced designs) between control and experimental arms as follows:

$$N_3^{(0)} = \frac{(1+1/\lambda) \{ (1 - \rho_1) + r_\tau N_1 \text{Var}_p(T) \} z_{\alpha, \varphi}^2}{N_2 N_1 \text{Var}_p(T) (\delta/\sigma)^2}, \quad (1)$$

where $T = 0, 1, \dots, N_1 - 1$; $\text{Var}_p(T)$ is the “population variance” of T ; σ^2 is the variance of outcome Y at baseline ($T = 0$); r_τ is the ratio of the variance of the random slope to σ^2 ($r_\tau = 0$ for fixed slope models); ρ_1 is the ratio of the sum of the variances of the second and third-level random intercepts to the variance of Y at baseline (for fixed-slope models, ρ_1 also corresponds to the correlations among repeated outcomes from the same subjects and is assumed to be constant over time [3]); δ is the XT interaction effect, i.e., the difference in mean slopes between intervention and control arms; and $z_{\alpha, \varphi} = \{ \Phi^{-1}(1-\alpha/2) + \Phi^{-1}(\varphi) \}$ with a two-sided significance level α and a desired power φ (see [3] for more details). Here,

$N_3^{(0)}$ = number of clusters for the control arm ($X=0$), N_2 = number of subjects per cluster, and N_1 = number of observations per subject. In equation (1), $N_3^{(0)}$ and N_2 are reciprocal to each other for the same power supporting the invariance over product property. Of note, $N_3^{(0)}$ and N_1 are not reciprocals since $\text{Var}_p(T)$ is a function of N_1 .

For trials with *randomization at the second level* with $1:\lambda$ unbalanced allocations between control and intervention arm subjects within clusters, the following formula essentially same as (1) can be applied owing to the uncorrelated slope estimates:

$$N_2^{(0)} = \frac{(1+1/\lambda) \{ (1 - \rho_1) + r_\tau N_1 \text{Var}_p(T) \} z_{\alpha, \varphi}^2}{N_3 N_1 \text{Var}_p(T) (\delta/\sigma)^2}. \quad (2)$$

. Here, N_3 = total number of clusters,

$N_2^{(0)}$ = number of subjects for a control arm per cluster ($X=0$), and N_1 is defined as before. Again, $N_2^{(0)}$ and N_3 are reciprocal to each other, preserving the invariance property.

Notably, the total number of subjects for fixed N_1 remains the same for trials with third (1) or second (2) level randomizations, i.e., $(1+\lambda)N_3^{(0)}N_2 = (1+\lambda)N_3N_2^{(0)}$. Simulation results (Table 1) validate equations (1) and (2), and the two aforementioned invariance properties.

Extension

For longitudinal-CRTs involving two experimental interventions X ($= 0$ for control and $= 1$ for experimental) and Z ($= 0$ for control and $= 1$ for experimental), it would be of interest to test whether the outcome trends (i.e., slopes) over the study period is beyond what would be expected if the effects of the two interventions on the slopes were additive. This hypothesis

can be tested in a 2×2 factorial longitudinal-CRT design setting by including a *three-way XZT interaction* between the two interventions and time in a linear mixed-effects linear model with fixed or random slopes for analysis of three-level data.

When randomization occurs at the third level, the clusters will be assigned to one of four (X, Z) combinations and all subjects within a cluster will uniformly receive an identical combination and then will be repeatedly evaluated for outcomes over time. Under this design, the number of clusters per (X, Z) cell required to detect the three-way *XZT* interaction effect is twice as large as that required per arm required to detect the two-way *XT* interaction effect. It follows that the required total number of observations or subjects will be four times larger. This proposition is based on a finding by Fleiss [7] that testing an interaction effect requires a sample size four times larger than required for testing a main effect in a 2×2 factorial cross-sectional design with one level data. Applications of the finding to cases with two level longitudinal data have been validated both theoretically and empirically with simulation studies [8, 9]. Further extension of the finding to an unbalanced longitudinal-CRT with third level randomization is straightforward, yielding the following formula:

$$N_3^{(0,0)} = \frac{(1 + 1/\lambda_{11} + 1/\lambda_{10} + 1/\lambda_{01}) \{ (1 - \rho_1) + r_\tau N_1 \text{Var}_p(T) \} z_{\alpha, \varphi}^2}{N_2 N_1 \text{Var}_p(T) (\delta/\sigma)^2}, \quad (3)$$

where $N_3^{(x,z)}$ is the number of clusters assigned to ($X=x, Z=z$) cell; $N_3^{(1,1)} = \lambda_{11} N_3^{(0,0)}$, $N_3^{(1,0)} = \lambda_{10} N_3^{(0,0)}$, $N_3^{(0,1)} = \lambda_{01} N_3^{(0,0)}$; N_2 = the number of subject per cluster; N_1 = number of observations per subject; and δ is the *XZT* interaction effect, i.e., the difference of differences in slope means. Equation (3) shows that the property of invariance over product is preserved since $N_3^{(0,0)}$ and N_2 are reciprocal to each other for the same power.

Furthermore, if $\lambda = 1$ in (1) and $\lambda_{01} = \lambda_{10} = \lambda_{11} = 1$ in (3), it can be seen that $N_3^{(0,0)} = 2N_3^{(0)}$ for the same effect sizes δ/σ as mentioned above.

When randomization occurs at the second level, the subjects within clusters will be randomly assigned to one of the four (X, Z) combinations. Since the slope estimates across the four (X, Z) combinations are uncorrelated regardless of level of randomization for the same reasons discussed in the appendix, the following formula similar to (3) can again be applied:

$$N_2^{(0,0)} = \frac{(1 + 1/\lambda_{11} + 1/\lambda_{10} + 1/\lambda_{01}) \{ (1 - \rho_1) + r_\tau N_1 \text{Var}_p(T) \} z_{\alpha, \varphi}^2}{N_3 N_1 \text{Var}_p(T) (\delta/\sigma)^2}, \quad (4)$$

where $N_2^{x,z}$ is the number of subjects assigned to ($X=x, Z=z$) cell within clusters; $N_2^{(1,1)} = \lambda_{11} N_2^{(0,0)}$, $N_2^{(1,0)} = \lambda_{10} N_2^{(0,0)}$, $N_2^{(0,1)} = \lambda_{01} N_2^{(0,0)}$; N_3 = the total number of clusters; N_1 = number of observations per subject. From equation (4): 1) the total number of observations or subjects required for testing the *XZT* interaction will be identical regardless of level of randomization since $N_3^{(0,0)} N_2 N_1 = N_3 N_2^{(0,0)} N_1$; and 2) the invariance over product is

preserved i.e., $N_2^{(0,0)}$ and N_3 are reciprocal. Under balanced designs, the number of subjects in each (X, Z) cell within clusters for testing the XZT interaction will be twice as large as that required in each group within clusters for testing the XT interaction, i.e., $N_2^{(0,0)} = 2N_2^{(0)}$ for the same δ/σ . Simulation results (Table 2) validate equations (3) and (4), the two invariance properties, and the aforementioned proposition.

Limitations

We considered only a normally-distributed continuous outcome Y . The applicability of the findings is not guaranteed for binary or other types of outcomes nor when the underlying assumptions such as normal random intercepts and slopes are violated.

Conclusion

The sample size approach for detecting a two-way or three-way interaction effect between time and interventions is the same regardless of whether randomization, balanced or unbalanced, occurs at the third or second level. This “invariance over level of randomization” property, along with the “invariance over product” property, offers clinical trialists and statisticians more flexibility in planning longitudinal clinical trials in regard to the level of randomization and allocation of clusters and subjects.

Acknowledgments

Funding: This study was supported in part by the NIH-funded Einstein-Montefiore Centers for AIDS research (P30 AI051519).

References

1. 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. *American Journal of Psychiatry*. 2005; 162:718–724. [PubMed: 15800144]
2. Heo M, Leon AC. Sample size requirements to detect an intervention by time interaction in longitudinal cluster randomized clinical trials. *Statistics in Medicine*. 2009; 28:1017–1027. [PubMed: 19153969]
3. Heo M, Xue X, Kim MY. Sample size requirements to detect an intervention by time interaction in longitudinal cluster randomized clinical trials with random slopes. *Comput Stat Data Anal*. 2013; 60:169–178. [PubMed: 23459110]
4. Moerbeek M. Randomization of clusters versus randomization of persons within clusters: Which is preferable? *American Statistician*. 2005; 59:173–179.
5. Raudenbush SW, Liu XF. Statistical power and optimal design for multisite randomized trials. *Psychological Methods*. 2000; 5:199–213. [PubMed: 10937329]
6. Moerbeek M, van Breukelen GJP, Berger MPF. Design issues for experiments in multilevel populations. *Journal of Educational and Behavioral Statistics*. 2000; 25:271–284.
7. Fleiss, JL. *The Design and Analysis of Clinical Experiments*. New York: Wiley & Sons; 1986.
8. Leon AC, Heo M. Sample sizes required to detect interactions between two binary fixed-effects in a mixed-effects linear regression model. *Computational Statistics & Data Analysis*. 2009; 53:603–608. [PubMed: 20084090]
9. Heo M, Leon AC. Sample sizes required to detect two-way and three-way interactions involving slope differences in mixed-effects linear models. *Journal of Biopharmaceutical Statistics*. 2010; 20:787–802. [PubMed: 20496206]

Appendix

Regardless of fixed or random slope model, it is sufficient to show that the slope estimates between groups within clusters are uncorrelated, that is to show that $Cov(\hat{\eta}_0, \hat{\eta}_1) = 0$, where

$$\hat{\eta}_0 = \sum_{i=1}^{N_3} \sum_{j(i)=1}^{N_2} \sum_{k=1}^{N_1} (T_k - \bar{T})(Y_{ij(i)k} - \bar{Y}_0) / N_3 N_2 N_1 \text{Var}_p(T)$$

is the slope estimate for a control

$$\hat{\eta}_1 = \sum_{i=1}^{N_3} \sum_{j'(i)=1}^{\lambda N_2} \sum_{k=1}^{N_1} (T_k - \bar{T})(Y_{ij'(i)k} - \bar{Y}_1) / N_3 N_2 N_1 \text{Var}_p(T)$$

for an experimental

arm. Please see [2] and [3] for detailed assumptions except that here the allocation ratio λ is included and the definitions of indices for within-cluster randomized longitudinal-CRTs are different as follows: the third level clusters are indexed by $i = 1, 2, \dots, N_3$; the second level subjects in the control arm within the i -th cluster are indexed by $j(i) = 1, 2, \dots, N_2 (= N_2^{(0)})$ and those in the experimental arm within the i -th cluster are indexed by $j'(i) = 1, 2, \dots, \lambda N_2$; and the first level outcome observations are indexed by $k = 1, 2, \dots, N_1$.

Let $W_k = (T_k - \bar{T})$, then it is sufficient to show that:

$$\text{Cov} \left(\sum_{i=1}^{N_3} \sum_{j(i)=1}^{N_2} \sum_{k=1}^{N_1} W_k Y_{ij(i)k}, \sum_{i=1}^{N_3} \sum_{j'(i)=1}^{\lambda N_2} \sum_{k=1}^{N_1} W_k Y_{ij'(i)k} \right) = \sum_{i=1}^{N_3} \text{Cov} \left(\sum_{j(i)=1}^{N_2} \sum_{k=1}^{N_1} W_k Y_{ij(i)k}, \sum_{j'(i)=1}^{\lambda N_2} \sum_{k=1}^{N_1} W_k Y_{ij'(i)k} \right) = 0.$$

. Observe that:

$$\begin{aligned} & \text{Cov} \left(\sum_{j(i)=1}^{N_2} \sum_{k=1}^{N_1} W_k Y_{ij(i)k}, \sum_{j'(i)=1}^{\lambda N_2} \sum_{k=1}^{N_1} W_k Y_{ij'(i)k} \right) \\ &= \sum_{j(i)=1}^{N_2} \sum_{j'(i)=1}^{\lambda N_2} \sum_{k=1}^{N_1} W_k^2 \text{Cov}(Y_{ij(i)k}, Y_{ij'(i)k}) + \sum_{j(i)=1}^{N_2} \sum_{j'(i)=1}^{\lambda N_2} \sum_{k=1}^{N_1} \sum_{k' \neq k}^{N_1} W_k W_{k'} \text{Cov}(Y_{ij(i)k}, Y_{ij'(i)k'}) \\ &= \sigma_3^2 \sum_{j(i)=1}^{N_2} \sum_{j'(i)=1}^{\lambda N_2} \left(\sum_{k=1}^{N_1} W_k^2 \right. \\ & \quad \left. + \sum_{k=1}^{N_1} \sum_{k' \neq k}^{N_1} W_k W_{k'} \right) = 0. \end{aligned}$$

. It is because that $\sum_{k=1}^{N_1} \sum_{k' \neq 1}^{N_1} W_k W_{k'} = - \sum_{k=1}^{N_1} W_k^2$ due to the fact that $\sum_{k' \neq 1}^{N_1} W_{k'} = -W_k$ since

$$\sum_{k=1}^{N_1} W_k = 0. \text{ It follows that } Cov(\hat{\eta}_0, \hat{\eta}_1) = 0.$$

Comparison of sample sizes, theoretical power and empirical power to detect a two-way interaction effect XT between intervention X and time T with the following parameters fixed: $\delta=0.125$, $\sigma^2=1$, $\rho_1=0.5$, $\rho_2=0.05$.

Table 1

Randomization at third level						
λ	r_e	$N_3^{(0)}$	N_2	N_1	Total N	$\tilde{\phi}$
1.0	0.0	4	15	5	600	0.865
		2	30	5	600	0.865
		0.1	12	15	5	1800
1.5	0.0	6	30	5	1800	0.865
		4	15	5	750	0.918
		2	30	5	750	0.918
1.5	0.1	10	15	5	1875	0.865
		5	30	5	1875	0.865

Randomization at second level						
λ	r_e	N_3	$N_2^{(0)}$	N_1	Total N	$\tilde{\phi}$
1.0	0.0	4	15	5	600	0.865
		2	30	5	600	0.865
		0.1	12	15	5	1800
1.5	0.0	6	30	5	1800	0.865
		4	15	5	750	0.918
		2	30	5	750	0.918
1.5	0.1	10	15	5	1875	0.865
		5	30	5	1875	0.865

Note: Total N = $(1+\lambda)N_3^{(0)}N_2N_1$ for the third-level randomization and $= (1+\lambda)N_3N_2^{(0)}N_1$ for the second-level randomization; ϕ = theoretical power, the inverse of equations (1) and (2); $\tilde{\phi}$ = empirical power based on 1000 simulations based on fitting of a pertinent mixed-effects linear model with SAS PROC MIXED; and ρ_2 = ratio of the variance of the third level random intercept to the variance of Y at baseline.

Table 2

Comparison of sample sizes, theoretical power and empirical power to detect a three-way interaction effect XZT between two interventions, X and Z , and time T with the following parameters fixed: $\delta = 0.125$, $\sigma^2 = 1$, $\rho_1 = 0.5$, $\rho_2 = 0.05$.

Randomization at third level						
$\lambda_{01}, \lambda_{10}, \lambda_{11}$	r_{τ}	$N_3^{(0,0)}$	N_2	N_1	Total N	$\tilde{\varphi}$
1, 1, 1	0.0	8	15	5	2400	0.855
		4	30	5	2400	0.863
		24	15	5	7200	0.865
1.5, 1, 1.5	0.1	12	30	5	7200	0.869
		8	15	5	3000	0.892
		4	30	5	3000	0.892
1.5, 1, 1.5	0.1	20	15	5	7500	0.866
		10	30	5	7500	0.831
		10	30	5	7500	0.879

Randomization at second level						
$\lambda_{01}, \lambda_{10}, \lambda_{11}$	r_{τ}	N_3	$N_2^{(0,0)}$	N_1	Total N	$\tilde{\varphi}$
1, 1, 1	0.0	8	15	5	2400	0.862
		4	30	5	2400	0.865
		24	15	5	7200	0.872
1.5, 1, 1.5	0.1	12	30	5	7200	0.876
		8	15	5	3000	0.892
		4	30	5	3000	0.892
1.5, 1, 1.5	0.1	20	15	5	7500	0.860
		10	30	5	7500	0.831
		10	30	5	7500	0.872

Note: Total N = $(1+\lambda_{01}+\lambda_{10}+\lambda_{11})N_3^{(0,0)}N_2N_1$ for the third-level randomization and $= (1+\lambda_{01}+\lambda_{10}+\lambda_{11})N_3N_2^{(0,0)}N_1$ for the second-level randomization; φ = theoretical power, inverse of equations (3) and (4); $\tilde{\varphi}$ = empirical power based on 1000 simulations based on fitting of a pertinent mixed-effects linear model with SAS PROC MIXED; and ρ_2 = ratio of the variance of the third level random intercept to the variance of Y at baseline.