

Design And Analysis of Longitudinal Multisite Randomized Trials: Estimation, Statistical Power, and Optimal Sample Size

Wei Li, University of Florida, wei.li@coe.ufl.edu
Spyros Konstantopoulos, Michigan State University, spyros@msu.edu
Zuchao Shen, University of Georgia, zuchao.shen@gmail.com

Background

Longitudinal multisite experimental designs are commonly employed in educational interventions, where, for example, students from the same schools are randomly assigned to either a treatment or control group and subsequently followed and measured over time. One objective of longitudinal studies is to examine how treatment effects evolve over time. Considering the typical nested data structure in multisite longitudinal studies (e.g., repeated measures nested within students nested within schools), multilevel models (MLM) with site random effects have been traditionally used to estimate the treatment effects on the linear or nonlinear rates of change (e.g., Raudenbush & Liu, 2001). Recent discussions on the design and analysis of multisite experimental studies (e.g., Miratrix, Weiss, & Henderson, 2021) suggested setting site effects fixed (e.g., site dummy variables) because of the concerns about potential correlations between treatment indicator and the site random effect. For example, the demand for a particular treatment might be higher for some schools than others, in which case the treatment is correlated with site characteristics represented by the site random effects in MLMs (Miratrix et al., 2021).

A crucial consideration in designing longitudinal experiments is determining the sample size allocation across levels and treatment conditions to ensure sufficient power to detect the effect of interest. Researchers typically plan their longitudinal studies with budget constraints in mind, as different sampling plans under the same budget can yield varying levels of statistical power. Prior studies have developed statistical power computation formulas to detect the main effects for three-level longitudinal multisite experiments using MLMs with random site effects (e.g., Li & Konstantopoulos, 2019). However, similar methods for the models with fixed site effects are lacking. Additionally, the optimal sample size computation methods under budget constraints for longitudinal multisite experiments have not been developed.

Purpose and Significance

This study contributes to the literature by providing methods of computing the statistical power and optimal sample sizes for longitudinal three-level multisite experiments (e.g., repeated measures, students, and schools) and implementing these methods into an R package to assist applied researchers in planning longitudinal experiments.

Research Design

Following prior literature in cross-sectional designs, this study considers three alternative methods to estimate the treatment effect on the linear rate of change: three-level growth model,

constant effect model, and fixed effect model. Specifically, the three-level random effect model is

Level 1 (repeated measures):

$$Y_{gij} = \alpha_{0ij} + \alpha_{1ij}c_{1g} + e_{gij}, e_{gij} \sim N(0, \sigma_e^2), \quad (1)$$

Level 2 (students):

$$\alpha_{0ij} = \beta_{00j} + \beta_{01j}T_{ij} + \mathbf{X}_{ij}\mathbf{B}_{01} + \xi_{A0ij}, \xi_{A0ij} \sim N(0, \tau_{0|T,\mathbf{X}}^2), \quad (2)$$

$$\alpha_{1ij} = \beta_{10j} + \beta_{11j}T_{ij} + \mathbf{X}_{ij}\mathbf{B}_{11} + \xi_{A1ij}, \xi_{A1ij} \sim N(0, \tau_{1|T,\mathbf{X}}^2), \quad (3)$$

Level 3 (schools):

$$\beta_{00j} = \gamma_{000} + \mathbf{Z}_j\mathbf{\Gamma}_{001} + u_{A00j}, u_{A00j} \sim N(0, \omega_{00|\mathbf{Z}}^2), \quad (4)$$

$$\beta_{01j} = \gamma_{010} + \mathbf{Z}_j\mathbf{\Gamma}_{011} + u_{A01j}, u_{A01j} \sim N(0, \omega_{10|\mathbf{Z}}^2), \quad (5)$$

$$\beta_{10j} = \gamma_{100} + \mathbf{Z}_j\mathbf{\Gamma}_{101} + u_{A10j}, u_{A10j} \sim N(0, \omega_{01|\mathbf{Z}}^2), \quad (6)$$

$$\beta_{11j} = \gamma_{110} + \mathbf{Z}_j\mathbf{\Gamma}_{111} + u_{A11j}, u_{A11j} \sim N(0, \omega_{11|\mathbf{Z}}^2), \quad (7)$$

where c_{1g} is a measure of time centered at the mid-timepoint, T_{ij} is a binary treatment indicator, $\mathbf{X}_{ij}$ is a vector of student-level covariates, $\mathbf{Z}_j$ is a vector of school-level covariates. The parameter of interest is γ_{110} , representing the treatment effect on the linear rate of change that is randomly varying among schools.

For the constant effect, the level-1 and level-2 models are the same as Equations 1-3, and the level-3 equations are:

Level 3 (schools):

$$\beta_{00j} = \gamma_{000} + \sum site_j, \quad (8)$$

$$\beta_{01j} = \gamma_{010}, \quad (9)$$

$$\beta_{10j} = \gamma_{100} + \sum site_j, \quad (10)$$

$$\beta_{11j} = \gamma_{110}, \quad (11)$$

where $site_j$ represents the site dummy variable that are grand centered. This model assumes the treatment effect on the linear rate of change is a constant (γ_{110}) across sites.

The fixed model also has the same level-1 and level-2 equations as Equations 1-3, and the level-3 equations become:

Level 3 (schools):

$$\beta_{00j} = \gamma_{000} + \sum site_j, \quad (12)$$

$$\beta_{01j} = \gamma_{010} + \sum site_j, \quad (13)$$

$$\beta_{10j} = \gamma_{100} + \sum site_j, \quad (14)$$

$$\beta_{11j} = \gamma_{110} + \sum site_j, \quad (15)$$

This model assumes the treatment effects non-randomly vary across sites and each site has its specific treatment effect.

Results:

Power Computation. We can test whether $\hat{\gamma}_{110} = 0$ using a t -test. Assuming the alternative hypothesis is true, the test statistic follows a non-central t -distribution, T' , with a non-centrality parameter:

$$\hat{\lambda} = \frac{\hat{\gamma}_{110}}{\sqrt{\text{Var}(\hat{\gamma}_{110})}}. \quad (16)$$

Under these specifications, the statistical power of a two-tailed test is (note $t_0 = t_{1-\frac{\alpha}{2}, df}$)

$$\text{Power} = 1 - P[T'(df, \hat{\lambda}) < t_0] + P[T'(df, \hat{\lambda}) \leq -t_0]. \quad (17)$$

where df is the degrees of freedom for the test.

Table 1 summarizes the formulas of $\text{Var}(\gamma_{110})$, the non-standardized non-centrality parameters and degrees of freedom (DF) for three models. We omit the procedure of standardizing the non-centrality parameters because of the word limit. As shown in Table 1, the non-centrality parameters for constant effect model and fixed effect model are the same, but their DFs are different. We will implement the formulas in Table 1 into a R package for applied researchers to plan longitudinal multisite experiments.

Optimal Sample Sizes. Suppose there are J level-3 units, n level-2 units within each level-3 units, and G repeated measures within each level-2. Let p represent the proportion of level-2 units in the treatment group. Then, the budget to collect data is

$$B = J[P(b_1^T nG + b_2^T n) + (1 - P)(b_1 nG + b_2 n) + b_3^T], \quad (18)$$

where the superscript T represents the treatment; b_1^T , b_2^T , b_3^T , b_1 , b_2 , and b_3 represent the budget to collect the data for one additional level-1, level-2, or level-3 unit in the treatment or control group, respectively. We can minimize $\text{Var}(\hat{\gamma}_{110})$ under a budget constrain (e.g., equation 18) with respect to J , G , n , and P , using the method of Lagrange Multiplier (Cochran, 1977) to compute the optimal sample size allocation:

$$L(J, n, G, P) = \text{Var}(\hat{\gamma}_{110}) + \Lambda(B - B_{cap}), \quad (19)$$

where Λ is the Lagrange Multiplier and B_{cap} represents the budget limit. The derivatives result in no-closed-form solutions, as presented by Shen & Kelcey (2020). However, we can further solve the numerical problems through the *uniroot* function in R along with iterations (Shen & Kelcey, 2020). Illustrative examples will be provided to demonstrate the applicability of the methods and tools in power and optimal sampling computations.

References:

- Li, W., & Konstantopoulos, S. (2019). Power computations for polynomial change in block randomized designs. *Journal of Experimental Education*, 87(4), 575-595.
- Miratrix, L. W., Weiss, M. J., & Henderson, B. (2021). An applied researcher's guide to estimating effects from multisite individually randomized trials: Estimands, estimators, and estimates. *Journal of Research on Educational Effectiveness*, 14(1), 270-308.
- Raudenbush, S. W., & Liu, X. F. (2001). Effects of study duration, frequency of observation, and sample size on power in studies of group differences in polynomial change. *Psychological Methods*, 6(4), 387-401.
- Shen, Z., & Kelcey, B. (2020). Optimal sample allocation under unequal costs in cluster-randomized trials. *Journal of Educational and Behavioral Statistics*, 45(4), 446-474.

Table 1. Summary of the Standardized Noncentrality Parameter, MDES, and Degrees of Freedom

Model	$Var(\gamma_{110})$	Non-Standardized Non-Centrality Parameter (λ)	Degree of Freedom
Three-level growth model	$Var(\gamma_{110}) = \frac{1}{Jn} (n\omega_{11 Z}^2 + \phi\tau_{1 T,X}^2 + \phi\sigma_1^2)$	$\lambda = \gamma_{110} \sqrt{\frac{Jn}{n\omega_{11 Z}^2 + \phi\tau_{1 T,X}^2 + \phi\sigma_1^2}}$	$J - g_3 - 1$
Constant effect model	$Var(\gamma_{110}) = \frac{\phi}{Jn} (\tau_{1 T,X}^2 + \sigma_1^2)$	$\lambda = \gamma_{110} \sqrt{\frac{Jn}{\phi\tau_{1 T,X}^2 + \phi\sigma_1^2}}$	$Jn - J - g_2 - 1$
Fixed effect model	$Var(\gamma_{110}) = \frac{\phi}{Jn} (\tau_{1 T,X}^2 + \sigma_1^2)$	$\lambda = \gamma_{110} \sqrt{\frac{Jn}{\phi\tau_{1 T,X}^2 + \phi\sigma_1^2}}$	$Jn - 2J - g_2 - 1$

Note: (1) $\phi = \frac{1}{p(1-p)}$. (2) $\sigma_1^2 = \frac{\sigma_e^2}{\sum c_{1g}^2}$. (3) g_3 represents the number of covariates at level 3 and g_2 represents the number of covariates at level 2.