Proceedings of the 89th Annual International Meeting of the Psychometric Society, Prague, Czech Republic (2024).

Revisiting the Estimation of Multilevel Modeling and Power Analysis for Multisite Randomized Trials

Zhenqiu Lu[0000-0001-9482-1368 *†]

† Quantitative Methodologies, University of Georgia, Athens, 30602, GA, USA
* Corresponding author. Email: zlu@uga.edu

Abstract
Challenges arise in multisite randomized trials in power analysis and optimal design due to the complexities of random coefficients and hierarchical structures. To address these challenges, this article revisits the estimation of multilevel models (MLM). It presents various approaches to deriving parameter estimates, incorporating different algebraic derivations—basic or matrix algebra, along with different designs—balanced and unbalanced designs, different coding schemes for treatment and control groups, and some specific special cases. By showing the similarities, differences, and relationships among these estimates, the study provides a clearer framework, enhances transparency, and improves interpretability. Furthermore, by comparing the derived results with existing literature, it establishes connections that deepen researchers' understanding of MLM estimation in study design and power analysis.

Keywords: multilevel model estimation, multisite randomized trials, balanced and unbalanced design, power analysis

1. Introduction

Power analysis (e.g., Cohen, 1988) is essential in randomized trials studies, particularly in determining the sample size to detect treatment effects. An optimal sample allocation (Raudenbush & Liu, 2000) is to maximize power while minimizing sampling costs. In power analysis and optimal design, the test statistic for the main effect of treatment follows a non-central t-distribution under the alternative hypothesis. Since power is positively related to the non-centrality parameter, maximizing power requires maximizing this parameter, which means minimizing the variance of the treatment effect estimate.

Despite its importance, several challenges exist. First, in multisite randomized trials, the incorporated random coefficients add complexity to estimation. Hierarchical linear models (HLM; e.g., Raudenbush & Bryk, 2002), also known as multilevel models (MLM; e.g., Snijders, & Bosker, 2012) or mixed-effect models (e.g., Fitzmaurice, Laird & Ware, 2012), have various estimation methods, such as full maximum likelihood estimation (FML; Goldstein, 1986; Longford, 1987), restricted/residual maximum likelihood estimation (RML; Mason et al., 1983; Raudenbush & Bryk, 1986), least square estimation (LS), exploratory estimation methods and others. Additionally, estimations can be derived using either basic algebra or matrix algebra, leading to discrepancies in formats. Second, most existing derivation in multisite randomized trials assume balanced designs (e.g., Raudenbush & Bryk, 2002; Dong, Kelcey & Spybrook, 2021), where control and experimental groups have equal sample sizes.

However, real-world studies often involve unbalanced designs, where sample sizes differ in control and experimental groups. Assuming balanced designs limits their applicability to practical scenarios. Third, different coding schemes for treatment and control groups create inconsistencies in estimation results. Some assign 0 to the control group and 1 to the experimental group, while others use -0.5 and +0.5, respectively. Many existing power analysis and optimal design results (e.g., Raudenbush & Liu, 2000; Dong, Kelcey & Spybrook, 2021) are derived under the latter coding scheme in balanced designs, limiting their generalizability to other coding schemes and unbalanced designs.

To address these challenges, this article aims to (1) provide a clear and systematic approach to deriving MLM parameter estimates and related results for mixed-effect models in multisite randomized trials to enhance clarity and interpretability, (2) extend estimation results to more general cases, particularly unbalanced designs, making it more applicable to real-world research, while also deriving balanced design results as a special case, (3) provide estimation approaches for different coding schemes to ensure transparency in their differences and implications, (4) compare the derived results with existing literature and establish connections among them, (5) help researchers to improve the understanding of theoretical foundation of MLM estimation in multisite randomized trials.

This paper is organized as follows. Section 2 reviews estimation in the general multilevel model using matrix algebra. Section 3 focuses on estimation in multisite randomized trials, using individual-level formulations and basic algebra. It also compares estimates derived from different coding schemes, study designs, and special cases. Section 4 applies the results to power analysis, and Section 5 concludes the paper.

2. The general MLM model

Suppose there are *J* sites and n_j is the sample size at the jth site. The total sample size N is $N = \Sigma n_j$. Let y_{ij} be the observed data point for object i at site j, and Y_j be an $n_j \times 1$ vector of variables y_{ij}. In a two-level hierarchical analysis, the general MLM has two levels (Raudenbush & Bryk, 2002). At Level 1, the model is

$$\boldsymbol{Y}_j = \boldsymbol{X}_j \boldsymbol{\beta}_j + \boldsymbol{\epsilon}_j \quad \text{with} \quad \boldsymbol{\epsilon}_j \sim N\left(\boldsymbol{0}, {\sigma_j}^2 \boldsymbol{I}\right) \tag{1}$$

where $\boldsymbol{X}_j$ is an $n_j \times (K+1)$ design matrix of independent variables at level 1, K is the number of independent variables, $\boldsymbol{\beta}_j$ is a $(K+1) \times 1$ vector of level 1 random parameters, I is an $n_j \times n_j$ identity matrix, and $\boldsymbol{\epsilon_j}$ is an $n_j \times 1$ vector of normally distributed random errors with mean **0** and covariance matrix ${\sigma_j}^2 \boldsymbol{I}$, which is usually assumed ${\sigma_j}^2 = \sigma^2$ for simplicity. At Level 2, the model is

$$\boldsymbol{\beta}_j = \boldsymbol{W}_j \boldsymbol{\gamma} + \boldsymbol{u}_j \quad \text{with} \quad \boldsymbol{u}_j \sim N(\boldsymbol{0}, \boldsymbol{\tau}) \tag{2}$$

where $\boldsymbol{W_j}$ is a $(K+1) \times m$ matrix of level 2 predictors, $\boldsymbol{\gamma}$ is an $m \times 1$ vector of fixed effects, $\boldsymbol{u_j}$ is a $(K+1) \times 1$ vector of level-2 random residuals assumed to follow a multivariate normal distribution with mean 0 and covariance matrix $\boldsymbol{\tau}$ of dimension $(K+1) \times (K+1)$. By combining (1) and (2), the composite form is

$$\boldsymbol{Y}_j = \boldsymbol{X}_j \boldsymbol{W}_j \boldsymbol{\gamma} + \boldsymbol{X}_j \boldsymbol{u}_j + \boldsymbol{\epsilon}_j$$

2. 1 The estimate of the random coefficient $\boldsymbol{\beta_j}$ at level 1

To estimate $\boldsymbol{\beta_j}$, we apply the ordinary least-squares (OLS) on regression and obtain:

$$\widehat{\boldsymbol{\beta}}_{\boldsymbol{j}} = \left(\boldsymbol{X}_{\boldsymbol{j}}' \boldsymbol{X}_{\boldsymbol{j}}\right)^{-\mathbf{1}} \boldsymbol{X}_{\boldsymbol{j}}' \boldsymbol{Y}_{\boldsymbol{j}} \tag{3}$$

2.1.1 Conditional expectation and conditional variance

The conditional expectation of $\widehat{\boldsymbol{\beta}}_{\boldsymbol{j}}$ is

$$E\left(\widehat{\boldsymbol{\beta}}_{\boldsymbol{j}} | \boldsymbol{\beta}_{\boldsymbol{j}}\right) = \left(\boldsymbol{X}_{\boldsymbol{j}}' \boldsymbol{X}_{\boldsymbol{j}}\right)^{-\mathbf{1}} \boldsymbol{X}_{\boldsymbol{j}}' \left(\boldsymbol{X}_{\boldsymbol{j}} \boldsymbol{\beta}_{\boldsymbol{j}}\right) = \boldsymbol{\beta}_{\boldsymbol{j}} \tag{4}$$

Assuming a known variance σ^2, the conditional variance of $\widehat{\boldsymbol{\beta}}_{\boldsymbol{j}}$ is

$$Var\left(\widehat{\boldsymbol{\beta}}_{\boldsymbol{j}} | \boldsymbol{\beta}_{\boldsymbol{j}}\right) = \sigma^2 \left(\boldsymbol{X}_{\boldsymbol{j}}' \boldsymbol{X}_{\boldsymbol{j}}\right)^{-1} = \boldsymbol{V}_{\boldsymbol{j}} \tag{5}$$

For simplicity, we use $\boldsymbol{V}_{\boldsymbol{j}}$ to denote $\sigma^2 \left(\boldsymbol{X}_{\boldsymbol{j}}' \boldsymbol{X}_{\boldsymbol{j}}\right)^{-1}$.

2.1.2 Unconditional expectation and unconditional variance

From equations (1), (2) and (3), we obtain

$$\widehat{\boldsymbol{\beta}}_{\boldsymbol{j}} = \boldsymbol{\beta}_j + \left(\boldsymbol{X}_{\boldsymbol{j}}' \boldsymbol{X}_{\boldsymbol{j}}\right)^{-\mathbf{1}} \boldsymbol{X}_{\boldsymbol{j}}' \boldsymbol{\epsilon}_j = \left(\boldsymbol{W}_j \boldsymbol{\gamma} + \boldsymbol{u}_j\right) + \boldsymbol{e}_j$$

where $\boldsymbol{e}_j = \left(\boldsymbol{X}_{\boldsymbol{j}}' \boldsymbol{X}_{\boldsymbol{j}}\right)^{-\mathbf{1}} \boldsymbol{X}_{\boldsymbol{j}}' \boldsymbol{\epsilon}_j$ with $\boldsymbol{e}_j \sim N\left(0, \boldsymbol{V}_j\right)$. Thus, the unconditional expectation of $\widehat{\boldsymbol{\beta}}_{\boldsymbol{j}}$ is

$$E\left(\widehat{\boldsymbol{\beta}}_{\boldsymbol{j}}\right) = E\left(\boldsymbol{\beta}_{\boldsymbol{j}}\right) = \boldsymbol{W}_{\boldsymbol{j}} \boldsymbol{\gamma} \tag{6}$$

and the unconditional variance of $\widehat{\boldsymbol{\beta}}_{\boldsymbol{j}}$, assuming known variance σ^2 and $\boldsymbol{\tau}$, is given by

$$Var\left(\widehat{\boldsymbol{\beta}}_j\right) = Var\left(\boldsymbol{u}_{\boldsymbol{j}} + \boldsymbol{e}_{\boldsymbol{j}}\right) = \boldsymbol{\tau} + \sigma^2 \left(\boldsymbol{X}_{\boldsymbol{j}}' \boldsymbol{X}_{\boldsymbol{j}}\right)^{-1} = \boldsymbol{\tau} + \boldsymbol{V}_j \tag{7}$$

2. 2 The estimate of the fixed effect $\boldsymbol{\gamma}$ at level 2

Since $\widehat{\boldsymbol{\beta}}_{\boldsymbol{j}} = \boldsymbol{W}_{\boldsymbol{j}} \boldsymbol{\gamma} + \boldsymbol{u}_{\boldsymbol{j}} + \boldsymbol{e}_{\boldsymbol{j}}$ where $\boldsymbol{e}_{\boldsymbol{j}} \sim N\left(0, \boldsymbol{V}_j\right)$ and $\boldsymbol{u}_{\boldsymbol{j}} \sim N(0, \boldsymbol{\tau})$, we apply the generalized least-squares (GLS) to estimate $\boldsymbol{\gamma}$ and obtain:

$$\widehat{\boldsymbol{\gamma}} = \left(\sum\nolimits_{j=1}^{J} \left(\boldsymbol{W}_{\boldsymbol{j}}' \left(\boldsymbol{\tau} + \boldsymbol{V}_{\boldsymbol{j}}\right)^{-\mathbf{1}} \boldsymbol{W}_{\boldsymbol{j}}\right)\right)^{-1} \sum\nolimits_{j=1}^{J} \left(\boldsymbol{W}_{\boldsymbol{j}}' \left(\boldsymbol{\tau} + \boldsymbol{V}_{\boldsymbol{j}}\right)^{-\mathbf{1}} \widehat{\boldsymbol{\beta}}_{\boldsymbol{j}}\right) \tag{8}$$

The expectation of $\widehat{\boldsymbol{\gamma}}$ is

$$E(\widehat{\boldsymbol{\gamma}}) = \boldsymbol{\gamma}$$

and the variance of $\widehat{\boldsymbol{\gamma}}$ is

$$Var(\widehat{\boldsymbol{\gamma}}) = \left(\sum\nolimits_{j=1}^{J} \left(\boldsymbol{W}_{\boldsymbol{j}}' \left(\boldsymbol{\tau} + \boldsymbol{V}_{\boldsymbol{j}}\right)^{-\mathbf{1}} \boldsymbol{W}_{\boldsymbol{j}}\right)\right)^{-1} \tag{9}$$

2. 3 The estimate of random components at both levels

When the residual variance σ^2 at level 1 and the residual covariance matrix τ at level 2 are unknown, there are no closed-form mathematical formulas available to estimate them. Instead, iterative numerical

procedures are used to obtain estimates. Several iterative estimation methods exist, including iterative GLS and iterative MLE.

For the iterative GLS, we first apply the OLS method to the composite model to obtain initial parameter estimates and an estimate of the residual covariance matrix. Then re-fit the composite model using GLS, treating the estimated residual covariance matrix as if it were the true residual covariance matrix. This yields updated estimates of the fixed effects parameters and residual covariance matrix. We repeat the previous step iteratively until the results converges.

For the iterative MLE, we first assign initial values to the unknown parameters. Then compute the likelihood function based on these parameter values, and update the parameter estimates based on the current likelihood function. We repeat the previous step multiple times until the difference between successive maximum likelihood values is less than a given criterion (e.g., 0.000001), indicating convergence. The final parameter estimates corresponding to the maximum likelihood function are the iterative ML estimates. There are two types of MLEs: FML and RML. While FML provides biased estimates of variance components, RML produces unbiased variance component estimates.

3. The mixed-effect model in multisite randomized trials

In multisite randomized trials, it is typically assumed that each site includes both control and experimental groups. Considering the simplest case with only a grouping variable X and no additional covariates, the Level 1 model is given by

$$y_{ij} = \beta_{0j} + \beta_{1j} X_{ij} + \epsilon_{ij} \quad \textit{with} \quad \epsilon_{ij} \sim i.i.d. N(0, \sigma^2) \tag{10}$$

assuming that both the control and experimental groups share the same residual variance σ^2. If $X_{ij} = 0$ for the control group (C) and $X_{ij} = 1$ for the experimental/treatment group (E), then β_{0j} represents the average outcome of the control group at site j, while β_{1j} represents the average difference between the treatment and control groups at site j. The Level 2 model is specified as:

$$\beta_{0j} = \gamma_{00} + u_{0j} \text{ and } \beta_{1j} = \gamma_{10} + u_{1j} \quad \text{with } \begin{pmatrix} u_{0j} \\ u_{1j} \end{pmatrix} \sim MN_2\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \tau_{00} & \tau_{01} \\ \tau_{10} & \tau_{11} \end{pmatrix}\right) \tag{11}$$

where γ_{00} is the mean of the control group across all *J* sites, and γ_{10} is the mean difference between the treatment and control group across *J* sites, $\tau_{00} = var(u_{0j}) = var(\beta_{0j})$ is the variance of control group means, $\tau_{11} = var(u_{1j}) = var(\beta_{1j})$ is the variance of mean difference between the treatment and control groups, and their covariance $\tau_{01} = \tau_{10} = cov(u_{0j}, u_{1j}) = var(\beta_{0j}, \beta_{1j})$. By combining (10) and (11), we obtain the composite model:

$$y_{ij} = (\gamma_{00} + \gamma_{10} X_{ij}) + (u_{0j} + u_{1j} X_{ij}) + \epsilon_{ij}$$

By comparing this model and the general MLM model, (1) - (2) reduce to (10) - (11), respectively, by setting $K = 1$, $m = 2$, and $\boldsymbol{W}_j = \boldsymbol{I}_2$.

Note that in multisite randomized trials, study designs can be balanced or unbalanced, depending on sample size allocation. Let n_j be the sample size at site j, with n_{jc} and n_{jE} the sample sizes of the control and experimental groups, respectively, such that $n_j = n_{jE} + n_{jc}$. In a balanced design, both groups have equal sample sizes, i.e., $n_j = 2n_{jE} = 2n_{jc}$. In contrast, an unbalanced design

allows the control and experiment groups to have different sample sizes. If further assume p as the proportion of participants assigned to the experimental group, then $n_{jE} = n_j p$ and $n_{jc} = n_j(1-p)$.

3. 1 The estimate of the random coefficient $\boldsymbol{\beta_{0j}}$ and $\boldsymbol{\beta_{1j}}$ at level 1

If we use dummy coding for the grouping variable *X* by setting $X_{ij} = 0$ for the control group and $X_{ij} = 1$ for the experimental group, then β_{0j} and β_{1j} can be estimated as

$$\hat{\beta}_{0j} = \bar{y}_{jc}$$

$$\hat{\beta}_{1j} = \bar{y}_{jE} - \bar{y}_{jc}$$

where $\bar{y}_{jc}$ and $\bar{y}_{jE}$ are the sample means of the control and experimental groups at site j, respectively. Since $\hat{\beta}_{0j}$ and $\hat{\beta}_{1j}$ are random-effect estimates, they have both conditional and unconditional expectations and variances. By comparing these estimates with (3), the coding of X matters. Different values of X lead to variations in the form of $\widehat{\boldsymbol{\beta}}_j$, especially for $\hat{\beta}_{0j}$, resulting in different interpretations.

3.1.1 Conditional expectation and conditional variance of the estimates

The conditional expectations of $\hat{\beta}_{0j}$ and $\hat{\beta}_{1j}$ are

$$E(\hat{\beta}_{oj}|\beta_{oj}) = \beta_{oj}$$

$$E(\hat{\beta}_{1j}|\beta_{1j}) = \beta_{1j}$$

By applying the central limit theorem (CLT), the conditional variances of $\hat{\beta}_{oj}$ and $\hat{\beta}_{1j}$ are

$$Var(\hat{\beta}_{oj}|\beta_{oj}) = \frac{\sigma^2}{n_{jc}} \tag{12}$$

$$Var(\hat{\beta}_{1j}|\beta_{1j}) = \sigma^2\left(\frac{1}{n_{jE}} + \frac{1}{n_{jc}}\right) \tag{13}$$

3.1.2 Unconditional expectation and unconditional variance of the estimates

By the law of total expectation $E(\mathrm{X}) = E\big(E(X|Y)\big)$, the unconditional expectations of $\hat{\beta}_{0j}$ and $\hat{\beta}_{1j}$ are

$$E(\hat{\beta}_{oj}) = E\left(E(\hat{\beta}_{oj}|\beta_{oj})\right) = E(\beta_{oj}) = \gamma_{00}$$

$$E(\hat{\beta}_{1j}) = E\left(E(\hat{\beta}_{1j}|\beta_{1j})\right) = E(\beta_{1j}) = \gamma_{10}$$

By the law of total variance $Var(\mathrm{X}) = Var\big(E(X|Y)\big) + E\big(Var(X|Y)\big)$, the unconditional variances of $\hat{\beta}_{0j}$ and $\hat{\beta}_{1j}$ are

$$\begin{aligned} Var(\hat{\beta}_{0j}) &= Var\big(E(\hat{\beta}_{0j}|\beta_{0j})\big) + E\big(Var(\hat{\beta}_{0j}|\beta_{0j})\big) \\ &= Var(\beta_{0j}) + E\big(Var(\hat{\beta}_{0j}|\beta_{0j})\big) = \tau_{00} + \frac{\sigma^2}{n_{jc}} \end{aligned} \tag{14}$$

$$\begin{aligned} Var(\hat{\beta}_{1j}) &= Var\big(E(\hat{\beta}_{1j}|\beta_{1j})\big) + E\big(Var(\hat{\beta}_{1j}|\beta_{1j})\big) \\ &= Var(\beta_{1j}) + E\big(Var(\hat{\beta}_{1j}|\beta_{1j})\big) = \tau_{11} + \sigma^2\left(\frac{1}{n_{jE}} + \frac{1}{n_{jc}}\right) \end{aligned} \tag{15}$$

In the special case where $n_j = n$ for all sites, the total sample size is $N = Jn$, with $n_E = np$ and $n_c = n(1-p)$. Under this condition, the variances in (14) and (15) simplify to

$$Var(\hat{\beta}_{0j}) = \tau_{00} + \frac{\sigma^2}{n(1-p)}$$

$$Var(\hat{\beta}_{1j}) = \tau_{11} + \sigma^2\left(\frac{1}{np} + \frac{1}{n(1-p)}\right)$$

3.1.3 Different coding schemes of X in a balanced design

Different codings of the group variable X lead to variations in the form of $\hat{\boldsymbol{\beta}}_j$ in equation (3), particularly $\hat{\beta}_{0j}$, resulting in different interpretations. In addition to dummy coding, another commonly used alternative is effect coding, where $X_{ij} = -0.5$ for the control group and $X_{ij} = +0.5$ for the experimental group (sometimes rescaled to -1/+1). This coding scheme is most appropriate in balanced designs, where the sample sizes in the control and experimental groups are equal (i.e., $n_{jE} = n_{jc} = \frac{1}{2}n_j$). In this case, the intercept β_{0j} represents the grand mean of *Y* at site j, and β_{1j} continues to represent the treatment effect at that site. When the design is unbalanced, the intercept becomes a weighted grand mean, making the interpretation of coefficients less intuitive. Therefore, under the $X_{ij} = -0.5/+0.5$ coding in a balanced design, the coefficients can be estimated as:

$$\hat{\beta}_{0j} = \bar{y}_j$$

$$\hat{\beta}_{1j} = \bar{y}_{jE} - \bar{y}_{jc}$$

Under this condition, the variances in (12) - (15) become:

$$Var(\hat{\beta}_{0j}|\beta_{0j}) = Var(\bar{y}_j) = \frac{\sigma^2}{n_j}$$

$$Var(\hat{\beta}_{1j}|\beta_{1j}) = Var(\bar{y}_{jE} - \bar{y}_{jc}) = \sigma^2\left(\frac{2}{n_j} + \frac{2}{n_j}\right) = \frac{4\sigma^2}{n_j}$$

$$Var(\hat{\beta}_{0j}) = Var(\beta_{0j}) + E\left(\frac{\sigma^2}{n_j}\right) = \tau_{00} + \frac{\sigma^2}{n_j} \tag{16}$$

$$Var(\hat{\beta}_{1j}) = Var(\beta_{1j}) + E\left(\frac{4\sigma^2}{n_j}\right) = \tau_{11} + \frac{4\sigma^2}{n_j} \tag{17}$$

In the special case where $n_j = n$ for all sites, these variances simplify to

$$Var(\hat{\beta}_{0j}|\beta_{0j}) = \frac{\sigma^2}{n} \tag{18}$$

$$Var(\hat{\beta}_{1j}|\beta_{1j}) = \frac{4\sigma^2}{n} \tag{19}$$

$$Var(\hat{\beta}_{0j}) = \tau_{00} + \frac{\sigma^2}{n} \tag{20}$$

$$Var(\hat{\beta}_{1j}) = \tau_{11} + \frac{4\sigma^2}{n} \tag{21}$$

The results in (18) - (19) are identical to those in equation (6) on page 201 of Raudenbush & Liu (2000), and (20) - (21) are identical to those in equation (7) on the same page. It is important to note that these results apply specifically to the coding scheme of $X_{ij} = -0.5\,/{+0.5}$ in a balanced design. When using $X_{ij} = 0/1$ coding, whether in a balanced or unbalanced design, the variances are given by (12)-(15).

3. 2 The estimate of the fixed-effect γ_{00} and γ_{10} at level 2

If $X_{ij} = 0$ for the control group and $X_{ij} = 1$ for the experimental group, then γ_{00} and γ_{10} can be estimated as

$$\hat{\gamma}_{00} = \bar{\hat{\beta}}_{0j} = \frac{1}{J}\sum_{j=1}^{J} \bar{y}_{jc} \tag{22}$$

$$\hat{\gamma}_{10} = \bar{\hat{\beta}}_{1j} = \frac{1}{J}\sum_{j=1}^{J} (\bar{y}_{jE} - \bar{y}_{jc}) \tag{23}$$

The expectation of the $\hat{\gamma}_{00}$ and $\hat{\gamma}_{10}$ are

$$E(\hat{\gamma}_{00}) = E\left(\bar{\hat{\beta}}_{0j}\right) = \gamma_{00}$$

$$E(\hat{\gamma}_{10}) = E\left(\bar{\hat{\beta}}_{1j}\right) = \gamma_{10}$$

By applying (14) and (15), the variances of $\hat{\gamma}_{00}$ and $\hat{\gamma}_{10}$ are

$$var(\hat{\gamma}_{00}) = var\left(\bar{\hat{\beta}}_{0j}\right) = var\left(\frac{1}{J}\sum_{j=1}^{J}\hat{\beta}_{0j}\right) = \frac{1}{J^2}\left(\sum_{j=1}^{J} var(\hat{\beta}_{0j})\right)$$

$$= \frac{1}{J^2}\left(\sum_{j=1}^{J}\left(\tau_{00} + \frac{\sigma^2}{n_{jc}}\right)\right) = \frac{\tau_{00}}{J} + \frac{\sigma^2}{J^2}\sum_{j=1}^{J}\frac{1}{n_{jc}} \tag{24}$$

$$var(\hat{\gamma}_{10}) = var\left(\bar{\hat{\beta}}_{1j}\right) = \frac{1}{J^2}\left(\sum_{j=1}^{J} var(\hat{\beta}_{1j})\right) = \frac{1}{J^2}\left(\sum_{j=1}^{J}\left(\tau_{11} + \sigma^2\left(\frac{1}{n_{jE}} + \frac{1}{n_{jc}}\right)\right)\right)$$

$$= \frac{\tau_{11}}{J} + \frac{\sigma^2}{J^2}\left(\sum_{j=1}^{J}\frac{1}{n_{jE}} + \sum_{j=1}^{J}\frac{1}{n_{jc}}\right) \tag{25}$$

In the special case where $n_j = n$ for all sites with $n_E = np$ and $n_c = n(1-p)$, equations (22) - (25) simplify to

$$\hat{\gamma}_{00} = \bar{y}_c$$

$$\hat{\gamma}_{10} = \bar{\hat{\beta}}_{1j} = \bar{y}_E - \bar{y}_c$$

$$var(\hat{\gamma}_{00}) = \frac{\tau_{00}}{J} + \frac{\sigma^2}{Jn(1-p)}$$

$$var(\hat{\gamma}_{10}) = \frac{\tau_{11}}{J} + \frac{\sigma^2}{J}\left(\frac{1}{np} + \frac{1}{n(1-p)}\right)$$

In a balanced design and $X_{ij} = -0.5$ for the control group and $X_{ij} = +0.5$ for the experimental group, γ_{00} is the expected value across all groups, while γ_{10} is the expected mean difference between the experimental and control groups. In this case, γ_{00} and γ_{10} can be estimated as follows:

$$\hat{\gamma}_{00} = \bar{\hat{\beta}}_{0j} = \frac{1}{J}\sum_{j=1}^{J}\bar{y}_j$$

$$\hat{\gamma}_{10} = \bar{\hat{\beta}}_{1j} = \frac{1}{J}\sum_{j=1}^{J}(\bar{y}_{jE} - \bar{y}_{jc})$$

By applying (16) and (17), the variances of the estimates $\hat{\gamma}_{00}$ and $\hat{\gamma}_{10}$ are

$$var(\hat{\gamma}_{00}) = var\left(\bar{\hat{\beta}}_{0j}\right) = \frac{1}{J^2}\sum_{j=1}^{J}\left(\tau_{00} + \frac{\sigma^2}{n_j}\right) = \frac{\tau_{00}}{J} + \frac{\sigma^2}{J^2}\sum_{j=1}^{J}\frac{1}{n_j} \tag{26}$$

$$var(\hat{\gamma}_{10}) = var\left(\bar{\hat{\beta}}_{1j}\right) = \frac{1}{J^2}\sum_{j=1}^{J}\left(\tau_{11} + \frac{4\sigma^2}{n_j}\right) = \frac{\tau_{11}}{J} + \frac{4\sigma^2}{J^2}\sum_{j=1}^{J}\frac{1}{n_j} \tag{27}$$

In the special case where $n_j = n$ for all sites, the estimates and their variances in equations (26)-(27) simplify to

$$\hat{\gamma}_{00} = \bar{y}$$

$$\hat{\gamma}_{10} = \bar{\hat{\beta}}_{1j} = \bar{y}_E - \bar{y}_c$$

$$var(\hat{\gamma}_{00}) = \frac{\tau_{00} + \sigma^2/n}{J} \tag{28}$$

$$var(\hat{\gamma}_{10}) = \frac{\tau_{11} + 4\sigma^2/n}{J} \tag{29}$$

The results (28) - (29) match those in equation (13) on page 202 of Raudenbush & Liu (2000). Again, these results specifically apply to the $X_{ij} = -0.5\ /+0.5$ coding in a balanced design. For the $X_{ij} = 0/1$ coding, whether in a balanced or unbalanced design, the variances are given by (24)-(25).

3. 3 The estimate of random components at both levels

When the level 1 random component σ^2 and the level 2 random component $\boldsymbol{\tau}$ are unknown, iterative algorithms are used for estimation. Suppose the Expectation-Maximization (EM) algorithm is applied. The E-step and M-step are detailed as follows.

E-step: The composite model can be expressed in matrix form as

$$\boldsymbol{Y}_j = \boldsymbol{X}_j\boldsymbol{\gamma} + \boldsymbol{X}_j\boldsymbol{u}_j + \boldsymbol{\epsilon}_j$$

where $\boldsymbol{Y}_j$ is an $n_j \times 1$ vector of y_{ij}, $\boldsymbol{X}_j$ is an $n_j \times 2$ design matrix, $\boldsymbol{\gamma}$ is a 2×1 vector of fixed-effect parameters, $\boldsymbol{u}_j$ is a 2×1 vector of level 2 randomness, and $\boldsymbol{\epsilon}_j$ is an $n_j \times 1$ vector of residues. Based on the assumptions in (1) and (2), the joint distribution of $\boldsymbol{Y}_j$ and $\boldsymbol{u}_j$ is given by

$$\begin{pmatrix} \boldsymbol{Y}_j \\ \boldsymbol{u}_j \end{pmatrix} \sim MN_{n_j+2}\left(\begin{pmatrix} \boldsymbol{X}_j\boldsymbol{\gamma} \\ \boldsymbol{0} \end{pmatrix}, \begin{pmatrix} \boldsymbol{X}_j\boldsymbol{\tau}\boldsymbol{X}_j' + \sigma^2\boldsymbol{I} & \boldsymbol{X}_j\boldsymbol{\tau} \\ \boldsymbol{\tau}\boldsymbol{X}_j{}' & \boldsymbol{\tau} \end{pmatrix}\right)$$

Therefore, the distribution of $\boldsymbol{u}_j$ conditional on $\boldsymbol{Y}_j$ is

$$(\boldsymbol{u}_j|\boldsymbol{Y}_j) \sim MN_2\left(\left(\boldsymbol{\tau}\boldsymbol{X}_j'(\boldsymbol{X}_j\boldsymbol{\tau}\boldsymbol{X}_j' + \sigma^2\boldsymbol{I})^{-1}(\boldsymbol{Y}_j - \boldsymbol{X}_j\boldsymbol{\gamma})\right), \left(\boldsymbol{\tau} - \boldsymbol{\tau}\boldsymbol{X}_j'(\boldsymbol{X}_j\boldsymbol{\tau}\boldsymbol{X}_j' + \sigma^2\boldsymbol{I})^{-1}\boldsymbol{X}_j\boldsymbol{\tau}\right)\right)$$

M-step: Once the random $\boldsymbol{u}_j$ are generated, the parameters $\boldsymbol{\gamma}$, σ^2 and $\boldsymbol{\tau}$ can be estimated by maximizing the likelihood or using least-square estimation methods.

$$\hat{\boldsymbol{\gamma}} = \left(\sum_{j=1}^{J} \boldsymbol{X}_j{}'\boldsymbol{X}_j\right)^{-1} \sum_{j=1}^{J} \boldsymbol{X}_j'(\boldsymbol{Y}_j - \boldsymbol{X}_j\boldsymbol{u}_j)$$

$$\hat{\sigma}^2 = \frac{1}{J}\sum_{j=1}^{J} \hat{\boldsymbol{\epsilon}}_j{}'\,\hat{\boldsymbol{\epsilon}}_j$$

$$\hat{\boldsymbol{\tau}} = \frac{1}{J-1}\sum_{j=1}^{J} \hat{\boldsymbol{u}}_j\hat{\boldsymbol{u}}_j{}'$$

Specifically, given the random coefficient estimates $\hat{\beta}_{0j}$ and $\hat{\beta}_{1j}$, the estimates $\hat{\tau}_{00}$ and $\hat{\tau}_{11}$ are

$$\hat{\tau}_{00} = \frac{1}{J-1}\sum_{j=1}^{J}\left(\hat{\beta}_{0j} - \hat{\gamma}_{00}\right)^2 - \frac{\hat{\sigma}^2}{n_{jc}} \tag{30}$$

$$\hat{\tau}_{11} = \frac{1}{J-1}\sum_{j=1}^{J}(\hat{\beta}_{1j} - \hat{\gamma}_{10})^2 - \hat{\sigma}^2\left(\frac{1}{n_{jE}} + \frac{1}{n_{jc}}\right) \tag{31}$$

where

$$\hat{\sigma}^2 = \frac{1}{J}\sum_{j=1}^{J}\frac{1}{n_j - 2}\left(\sum_{i=1}^{n_j}\left(y_{ij} - \hat{\beta}_{0j} - \hat{\beta}_{1j}\,X_{ij}\right)^2\right)$$

The results (30)-(31) are derived from (14)-(15) and

$$Var(\hat{\beta}_{0j}) = \frac{1}{J-1}\sum_{j=1}^{J}(\hat{\beta}_{0j} - \hat{\gamma}_{00})^2 = \hat{\tau}_{00} + \frac{\hat{\sigma}^2}{n_{jc}}$$

$$Var(\hat{\beta}_{1j}) = \frac{1}{J-1}\sum_{j=1}^{J}(\hat{\beta}_{1j} - \hat{\gamma}_{10})^2 = \hat{\tau}_{11} + \hat{\sigma}^2\left(\frac{1}{n_{jE}} + \frac{1}{n_{jc}}\right)$$

When the design is balanced, if we assume $n_j = n$ and the coding is $X_{ij} = -0.5\ /+0.5$, then (30) - (31) simply to

$$\hat{\tau}_{00} = \frac{1}{J-1}\sum_{j=1}^{J}\left(\hat{\beta}_{0j} - \hat{\gamma}_{00}\right)^2 - \frac{\hat{\sigma}^2}{n} \tag{32}$$

$$\hat{\tau}_{11} = \frac{1}{J-1}\sum_{j=1}^{J}(\hat{\beta}_{1j} - \hat{\gamma}_{10})^2 - \frac{4\hat{\sigma}^2}{n} \tag{33}$$

The results (32) - (33) align with Equation (12) on page 202 of Raudenbush & Liu (2000).

4. Power analysis

With the derived parameter estimates, now we can compute the statistical power for testing the average treatment effect. In power analysis, the t statistic for the main effect of treatment γ_{10} with degrees of freedom *J-1* is calculated as follows:

$$t_{J-1} = \frac{\hat{\gamma}_{10}}{\sqrt{\widehat{var}\ (\hat{\gamma}_{10})}}$$

For a non-zero treatment effect H_a under the alternative hypothesis, the *t* statistic follows a non-central *t* distribution with the degrees of freedom *J-1* and a non-centrality parameter λ_{nc}.

In an unbalanced design,

$$\lambda_{nc} = \frac{\gamma_{10}}{\sqrt{\frac{\tau_{11}}{J} + \frac{\sigma^2}{J^2}\left(\sum_{j=1}^{J}\frac{1}{n_{jE}} + \sum_{j=1}^{J}\frac{1}{n_{jc}}\right)}}$$

If the type-1 error is α, then the critical value for a two-sided *t* test is $t_0 = t(1-\alpha/2, J-1)$, and the power $= 1 - \Phi\left(t_0, J-1,\ \lambda_{nc}\right) + \Phi(-t_o, J-1, \lambda_{nc})$, and in a one-sided t test, power $= 1 - \Phi\left(t_0, J-1,\ \lambda_{nc}\right)$, where $\Phi\left(q, df,\ \lambda_{nc}\right)$ is the cumulative distribution function (CDF) of a non-central *t* distribution.

In a balanced design, the λ_{nc} simplify to

$$\lambda_{nc} = \frac{\bar{y}_E - \bar{y}_c}{\sqrt{\frac{\tau_{11}}{J} + \frac{4\sigma^2}{J^2}\sum_{j=1}^{J}\frac{1}{n_j}}}$$

In a two-sided *t* test, power $= 1 - \Phi\left(t_0, J-1,\ \lambda_{nc}\right) + \Phi(-t_o, J-1, \lambda_{nc})$, and in a one-sided *t* test, power $= 1 - \Phi\left(t_0, J-1,\ \lambda_{nc}\right)$. In the special case where $n_j = n$ for all sites, the non-centrality parameter λ_{nc} becomes

$$\lambda_{nc} = \frac{\bar{y}_E - \bar{y}_c}{\sqrt{\frac{\tau_{11} + 4\sigma^2/n}{J}}}$$

This λ_{nc} corresponds to the square root of the λ in Equation (15) on page 202 of Raudenbush & Liu (2000), where they used a non-central *F* distribution instead of a non-central *t* distribution. Their λ parameter was derived under the assumption of $n_j = n$ in a balanced design, which is a special case of the more general scenarios developed in this article.

5. Discussion

This article revisits the estimation of multilevel models in power analysis and optimal design, with a focus on mixed-effect models in multisite randomized trials. It presents various approaches to deriving parameter estimates by incorporating matrix and individual forms, balanced and unbalanced designs, different coding schemes, and some special cases. By showing the similarities, differences, and relationships among these estimates, the study provides a clearer framework, enhances transparency, and improves interpretability. This work can be extended to other designs in power analysis and optimal design, as well as broader research contexts. Future studies can further explore their applications across diverse study designs and practical settings.

Acknowledgement

Funding Statement: None

References

Cohen, J. (1988). Statistical power analysis for the behavioral sciences (2nd ed.). Lawrence Erlbaum Associates. https://doi.org/10.4324/9780203771587

Dong, N., Kelcey, B., & Spybrook, J. (2021). Design considerations in multisite randomized trials probing moderated treatment effects. Journal of Educational and Behavioral Statistics, 46(5), 527-559. https://doi.org/10.3102/1076998620961492

Fitzmaurice, G. M., Laird, N. M., & Ware, J. H. (2012). Applied longitudinal analysis. John Wiley & Sons. https://DOI.org/10.1002/9781119513469

Goldstein, H. (1986). Multilevel mixed linear model analysis using iterative generalized least squares. Biometrika, 73(1), 43-56. https://doi.org/10.1093/biomet/73.1.43

Longford, N. T. (1987). A fast scoring algorithm for maximum likelihood estimation in unbalanced mixed models with nested random effects. Biometrika, 74(4), 817-827. https://doi.org/10.1093/biomet/74.4.817

Mason, W. M., Wong, G. Y., & Entwisle, B. (1983). Contextual analysis through the multilevel linear model. Sociological methodology, 14, 72-103. https://doi.org/10.2307/270903

Raudenbush, S., & Bryk, A. S. (1986). A hierarchical model for studying school effects. Sociology of education, 59(1). 1-17. https://doi.org/10.2307/2112482

Raudenbush, S. W., & Liu, X. (2000). Statistical power and optimal design for multisite randomized trials. Psychological methods, 5(2), 199-213. https://doi.org/10.1037/1082-989X.5.2.199

Raudenbush, S. W. & Bryk, A. S. (2002). Hierarchical linear models: Applications and data analysis methods. Advanced Quantitative Techniques in the Social Sciences Series/SAGE. https://psycnet.apa.org/record/1992-97494-000

Snijders, T. A. B., & Bosker, R. J. (2012). Multilevel analysis: An introduction to basic and advanced multilevel modeling (2nd ed.). Sage. https://uk.sagepub.com/en-gb/eur/multilevel-analysis/book234191