Abstract Title Page

Title: Power and Coverage of Partially Nested Randomized Trials

Authors and Affiliations: Ariel M Aloe, Zhijiang 'Van' Liu, & Hyeri Hong; The University of Iowa

Background/Context:
Partially nested randomized trials are special cases of cluster randomized trials (Lohr, Schochet, & Sanders, 2014). The main distinction is that while in cluster randomized trials both treatment and control possess cluster structures, only the treatment group possess a cluster structure in partially nested randomized trials. Partially nested structure is common in education interventions. The most obvious example is the evaluation of summer interventions. For example, while evaluating summer reading interventions, the treatment group is clustered with interventionists while the control group is not (e.g., Reed, Aloe, Reeger, & Folsom, 2019). In the last decade several manuscripts considered issues related to partially nested design. For example, Lohr et al. (2014) discussed the design and analyses of partially nested randomized trials, Bauer, Sterba, and Hallfors (2008) considered issues in the analyses of partially nested randomized trials, while Hedges and Citkowicz (2015) and Lai and Kwok (2016) considered the estimation of effect sizes (i.e., the standardized mean differences).

However, none of the current literature in education explored directly issues related to the power of cluster randomized trials. The power of the statistical model plays a direct role in the design stage of any randomized trial. Determining the number of units needed to implement an experimental design is required by most (if not all) funding agencies. However, most statistical softwares used by educational researchers at the planning stage of their studies do not allow for the direct estimation of power for partially nested randomized trials. Moreover, within the popular mixed models packages in `R` language such as `lme4`, researchers do not have the flexibility to estimate a partially nested model properly directly. It appears that some researchers intending to account for the partially nested data structure on `lme4` attempt to achieve their goal by coding the control group as nested within one single ID. We will also demonstrate that this practice does not produce an optimal result.

Purpose / Objective / Research Question / Focus of Study:
We developed a new package in `R` names `pcluster` that allows the proper estimation of partially nested models. We investigate via Monte Carlo Simulations the performance of implementing a fully nested model with all members of the control group belonging to the same group against the proper implementation of the partially nested model. The details of both models are discussed in the next section.

The goal of our first simulation was to examine the performance of partially nested model against mixed model (treating all control individuals as one group) in the following three aspects:

- Empirical sizes and powers of testing $H_0 : \mu_{\cdot}^{T} - \mu^{C} = 0$ using 95% CIs
- Coverage rates of 80% CIs of $\mu_{\cdot}^{T} - \mu^{C}$, and
- Coverage rates of 95% CIs of δ

The goal of our second simulation was to compare the performance of CIs under three models:

- OLS model (ignore clustering of individuals, i.e., $\rho = 0$),
- mixed model (treating all control individuals as one group),
- partially-clustered model.

Significance / Novelty of study:
R package `pcluster` is created to accommodate the special structure of (weighted) partially-clustered model given the inflexibility of currently available packages for mixed model analysis.

Statistical, Measurement, or Econometric Model:
A regular mixed model framework in matrix form is

$$\underline{\mathbf{y}} = \mathbf{X}\underline{\boldsymbol{\beta}} + \mathbf{Z}\underline{\mathbf{b}} + \underline{\boldsymbol{\varepsilon}}, \tag{1}$$

where $\underline{\mathbf{y}}$ is the vector of responses, $\mathbf{X}$ and $\mathbf{Z}$ are known design matrices for fixed- and random-effects, $\underline{\boldsymbol{\beta}}$ and $\underline{\mathbf{b}}$ are unknown vectors of fixed- and random-effects and $\underline{\boldsymbol{\varepsilon}}$ is the vector of errors. Moreover,

$$\begin{bmatrix} \underline{\mathbf{b}} \\ \underline{\boldsymbol{\varepsilon}} \end{bmatrix} \sim N\left(\underline{\mathbf{0}}, \begin{bmatrix} \tau^2\mathbf{I} & \mathbf{0} \\ \mathbf{0} & \sigma^2\mathbf{I} \end{bmatrix}\right).$$

In practice, primary study researchers might encounter subjects with multiple membership. For instance, for a study of a specific learning technique on children's verbal skills, children might be clustered according to different schools, classes and teachers. The mixed model framework with q memberships, in matrix form, is

$$\underline{\mathbf{y}} = \mathbf{X}\underline{\boldsymbol{\beta}} + \sum_{j=1}^{q} \mathbf{Z}_j\underline{\mathbf{b}}_j + \underline{\boldsymbol{\varepsilon}}, \tag{2}$$

where $\underline{\mathbf{y}}$ is the vector of responses, $\mathbf{X}$ and $\mathbf{Z}_j$'s are known design matrices for fixed- and random-effects, $\underline{\boldsymbol{\beta}}$ and $\underline{\mathbf{b}}_j$'s are unknown vectors of fixed- and random-effects and $\underline{\boldsymbol{\varepsilon}}$ is the vector of errors. Moreover,

$$\begin{bmatrix} \underline{\mathbf{b}}_1 \\ \vdots \\ \underline{\mathbf{b}}_q \\ \underline{\boldsymbol{\varepsilon}} \end{bmatrix} \sim N\left(\underline{\mathbf{0}}, \begin{bmatrix} \tau_1^2\mathbf{I} & & & \\ & \ddots & & \\ & & \tau_q^2\mathbf{I} & \\ & & & \sigma^2\mathbf{I} \end{bmatrix}\right).$$

It can be easily seen that (1) is simply a special case of (2) when $q = 1$. The parameters can be estimated with MLE, based on regular or REML log-likelihood. The regular log-likelihood function is

$$\ell = -\frac{1}{2}\log\left(|\boldsymbol{\Sigma}|\right) - \frac{1}{2}\left(\underline{\mathbf{y}} - \mathbf{X}\underline{\boldsymbol{\beta}}\right)'\boldsymbol{\Sigma}^{-1}\left(\underline{\mathbf{y}} - \mathbf{X}\underline{\boldsymbol{\beta}}\right) - \frac{n}{2}\log\left(2\pi\right), \tag{3}$$

where

$$\boldsymbol{\Sigma} = \sum_{j=1}^{q} \tau_j^2\mathbf{Z}_j\mathbf{Z}_j' + \sigma^2\mathbf{I}. \tag{4}$$

The REML log-likelihood function is

$$\ell_R = -\frac{1}{2}\log\left(|\mathbf{A}'\mathbf{\Sigma}\mathbf{A}|\right) - \frac{1}{2}\underline{\mathbf{y}}'\mathbf{A}'(\mathbf{A}'\mathbf{\Sigma}\mathbf{A})^{-1}\mathbf{A}\underline{\mathbf{y}} - \frac{n-r}{2}\log(2\pi), \tag{5}$$

where $r = n - \text{rank}(\mathbf{X})$ and $\mathbf{A}$ is column combination of $n-r$ linearly independent vectors $\underline{\mathbf{a}}_i$ such that $\underline{\mathbf{a}}_i'\mathbf{X} = \underline{\mathbf{0}}$ for $i = 1, 2, ..., n-r$. In `pcluster` package, $\underline{\mathbf{a}}_i$ are chosen from $\mathbf{I} - \mathbf{P}_\mathbf{X} = \mathbf{I} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$. Alternatively, PROC MIXED in SAS uses the following REML log-likelihood

$$\ell_{R2} = -\frac{1}{2}\log(|\mathbf{\Sigma}|) - \frac{1}{2}\log\left(|\mathbf{X}'\mathbf{\Sigma}^{-1}\mathbf{X}|\right) - \frac{1}{2}\underline{\mathbf{r}}'\mathbf{\Sigma}^{-1}\underline{\mathbf{r}} - \frac{n-r}{2}\log(2\pi), \tag{6}$$

where $\underline{\mathbf{r}} = \underline{\mathbf{y}} - \mathbf{X}(\mathbf{X}'\mathbf{\Sigma}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{\Sigma}^{-1}\underline{\mathbf{y}}$. Once the estimates of variance components, i.e., $\widehat{\tau}_j^2$'s and $\widehat{\sigma}^2$, are obtained from maximizing ℓ_{R1} or ℓ_{R2}, the estimate of fixed effect is

$$\underline{\widehat{\boldsymbol{\beta}}} = \left(\mathbf{X}'\widehat{\mathbf{\Sigma}}^{-1}\mathbf{X}\right)^{-1}\mathbf{X}'\widehat{\mathbf{\Sigma}}^{-1}\underline{\mathbf{y}}, \tag{7}$$

where

$$\widehat{\mathbf{\Sigma}} = \sum_{j=1}^{q}\widehat{\tau}_j^2\mathbf{Z}_j\mathbf{Z}_j' + \widehat{\sigma}^2\mathbf{I}. \tag{8}$$

We conducted two Monte Carlo simulation studies. For each condition we performed a 1,000 replications using the `lme4` and the `pcluster` packages.

For this simulation, we choose $\mu^C = 0$, $\sigma_W^2 = \tau^2 = 1$, hence ICC $\rho = \frac{\tau^2}{\sigma_W^2+\tau^2} = 0.5$. $n_1 = \cdots = n_m = n$, $N_C = N_T = mn$, then $N = N_C + N_T = 2mn$. Three varying quantities are cluster number $m = (10, 20, 30)$, cluster size $n = (5, 10, 15, 20, 25)$ and $d = (0, 0.25, 0.5, 0.75)$. Since $d = \frac{\mu_\cdot^T - \mu^C}{\sqrt{\sigma_W^2+\tau^2}}$, d corresponds to $\mu_\cdot^T = (0, \sqrt{2}/4, \sqrt{2}/2, 3\sqrt{2}/4)$.

For this second simulation, we choose $\mu^C = 0$, $\sigma_W^2 = 1$. $n_1 = \cdots = n_m = n$, $N_C = N_T = mn$, then $N = N_C + N_T = 2mn$. Four varying quantities are cluster number $m = (10, 20, 30)$, cluster size $n = (5, 10, 15, 20, 25)$, $d = (0, 0.25, 0.50, 0.75)$ and $\tau^2 = (0.10, 0.25, 0.50, 0.75)$. Thus, the corresponding $\rho = (0.091, 0.200, 0.333, 0.500)$. Since $d = \frac{\mu_\cdot^T - \mu^C}{\sqrt{\sigma_W^2+\tau^2}}$, d corresponds to $\mu_\cdot^T = (0, \sqrt{2}/4, \sqrt{2}/2, 3\sqrt{2}/4)$.

Usefulness / Applicability of Method: The results of our two Monte Carlo simulation studies demonstrate the importance of using a package that specifies the model properly. The beta version of the `pcluster` package will be available to researchers by the time of this presentation via `GitHub` and shortly after deposited in `CRAN`. The models are specified very similarly to the popular `lme4` package. The `pcluster` package can be used at the planning stage of a study to determine the number of units needed via simulation. Certainly, it can be used to properly implement the analyses of partially nested models. Among its functionality, the `pcluster` package allows researchers to specify models with common or different within variance components, different levels of nested structure between treatment and control groups, and user defined weights. Thus, this presentation not only demonstrates the importance of properly specified partially nested models, but it also will introduce a `R` package for researchers to be able to implement these models in their own research.

Conclusions:
From the results of our two simulation studies, we can conclude that, compared to partially-clustered model, mixed model produces severely conservative results, while OLS model has quite radical ones. Simulation I indicates that the empirical sizes of the Type I error under partially-clustered model are close to nominal level of 0.05, especially when the cluster count m increases, while the empirical sizes under mixed model are extremely low. The coverage rates under partially-clustered model start to be slightly below the nominal level of 0.80 when the cluster count m is small, and fluctuate around the nominal level when $m = 30$. However, under the mixed model setting, the coverage rates are extremely high (almost 100% in many cases). These high coverage rates under mixed model are at the cost of precision. On average, the width of CI from mixed model is roughly 4 times of that from partially-clustered model, which is the result of huge standard error of $\widehat{\mu}_{\cdot}^{T} - \widehat{\mu}^{C}$ from the mixed model.

Results of our second simulation indicated that the differences among three models are small when ρ is small, when the covariance structure is close to the OLS setting. The empirical sizes under mixed model are extremely low, which are almost 0 when ICC ρ and the cluster count m are large. For OLS model, the empirical sizes are always beyond nominal level of 0.05, and the inflation is more severe when the cluster size n or ρ increases, but seems consistent when the cluster count m changes. Generally, the empirical power values under OLS model are always higher than those under the partially nested model, while those under the mixed model are even below those under partially nested model. The empirical power levels under mixed model are somewhat consistent against varying cluster count m, but they either stay at a lower level than the empirical powers under other two models, when ρ is small and δ is large, or just stay at zero when ρ is large. The empirical powers of both OLS and partially nested models have increasing patterns as the cluster count m and cluster size n increase, when δ is small, and they stay around 1 when either cluster count m and δ are large.

References

Bauer, D. J., Sterba, S. K., & Hallfors, D. D. (2008). Evaluating group-based interventions when control participants are ungrouped. *Multivariate behavioral research*, *43*(2), 210–236.

Hedges, L. V. & Citkowicz, M. (2015). Estimating effect size when there is clustering in one treatment group. *Behavior research methods*, *47*(4), 1295–1308.

Lai, M. H. & Kwok, O.-m. (2016). Estimating standardized effect sizes for two-and three-level partially nested data. *Multivariate behavioral research*, *51*(6), 740–756.

Lohr, S., Schochet, P. Z., & Sanders, E. (2014). Partially nested randomized controlled trials in education research: a guide to design and analysis. ncer 2014-2000. *National Center for Education Research*.

Reed, D. K., Aloe, A. M., Reeger, A. J., & Folsom, J. S. (2019). Defining summer gain among elementary students with or at risk for reading disabilities. *Exceptional Children*, *85*(4), 413–431.