Power estimation methods

Porter, Miratrix, Hunter

May 28, 2021

Contents

1 Introduction

1.1 Model taxonomy

A basic note on notation: we use i to index individuals, j to index schools, k to index districts, and m to index outcomes.

We create a taxonomy of these models in addition to the categorization by clustering, blocking, etc. These models vary in the following ways:

- Whether the school-level and district-level intercepts are:
  - fixed ($u_{0,jkm}$ and $w_{0,jkm}$ are fixed effects constrained to have mean 0)
  - random ($u_{0,jkm}$ and $w_{0,jkm}$ are considered to be Normally distributed, allowing for partial pooling)
- Whether the school-level and district-level treatment effects/slopes are:
  - constant, e.g. all schools and districts are modeled as having a single average impact ($u_{1,jkm} = 0$ and $w_{1,km} = 0$)
  - fixed, e.g. each school or district has an individual estimated impact ($u_{1,jkm}$ are fixed values constrained to have mean 0)
  - random ($u_{1,jkm}$ and $w_{1,km}$ are Normally distributed around a mean impact)
- Whether school covariates X_{jkmr} or district covariates V_{kms} are included in the model for the intercept. Note that in some models, PowerUp incorporates covariates into the slope, allowing for, in principle, heterogeneous treatment efects correlated to said covariates. However, we never make this assumption, as it is unlikely to help with an evaluationg targeting an overall average impact. Including covariates in the slope terms allows for modeling treatment effect heterogeneity in principle, but this can add complexity with estimation. When this difference occurs, it is noted.

1.2 Proposed naming conventions

Proposed names would be: levels_treatment_intercepts_effects

Examples:

- 2level_b2_ir_er: 2 level, blocked on the second level, random intercept, random treatment effect. Corresponds to current blocked_i1_2r.
- 3level_b3c2_if_ec: 3 level, blocked on third level, clustered on second level, fixed intercepts, constant treatment effect. Corresponds to current blocked_c2_3f.

1.3 Notation

Tables 1-4 outline all important quantities for calculating power, MDES, and sample size. Table 5 provides a reference for translating notation between this document and PowerUpR! We now provide further clarification on the parameters in Table 4; for a more detailed discussion of these expressions, see Section 8.

The quantity ICC is the Intraclass Correlation for the unconditional model.

$$
\begin{aligned}
\mathrm{ICC}_{3,m} &= \frac{Var(\mu_{0,km})}{Var(Y_{ijkm}(0))} = \frac{\xi_m^2 + \eta_{0,m}^2}{\xi_m^2 + \eta_{0,m}^2 + \delta_m^2 + \tau_{0,m}^2 + \gamma_m^2 + \sigma_m^2} \\
\mathrm{ICC}_{2,m} &= \frac{Var(\theta_{0,jkm} \mid \mu_{0,km})}{Var(Y_{ijkm}(0))} = \frac{\delta_m^2 + \tau_{0,m}^2}{\xi_m^2 + \eta_{0,m}^2 + \delta_m^2 + \tau_{0,m}^2 + \gamma_m^2 + \sigma_m^2}
\end{aligned}
$$

The quantity ω is the ratio between impact variation and mean variation.

$$
\begin{aligned}
\omega_{3,m} &= \frac{Var(\mu_{1,jkm})}{Var(\mu_{0,km})} = \frac{\eta_{1,m}^2}{\left(\xi_m^2 + \eta_{0,m}^2\right)} \\
\omega_{2,m} &= \frac{Var(\psi_{1,jkm} \mid \mu_{1,km})}{Var(\theta_{0,jkm} \mid \mu_{0,km})} = \frac{\tau_{1,m}^2}{\left(\delta_m^2 + \tau_{0,m}^2\right)}
\end{aligned}
$$

The R^2 expressions are the percent of variation at a particular level predicted by covariates specific to that level.

$$
\begin{aligned}
R_{3,m}^2 &= 1 - \frac{Var(w_{0,km})}{Var(\mu_{0,km})} \\
R_{2,m}^2 &= 1 - \frac{Var(u_{0,jkm})}{Var(\theta_{0,jkm} \mid D_{id})} \\
R_{1,m}^2 &= 1 - \frac{Var(r_{ijkm})}{Var(Y_{ijkm}(0) \mid S_{id}, D_{id})}
\end{aligned}
$$

Param	Description
M	Number of outcomes
J	Number of schools in each district (assumed constant across districts)
K	Number of districts
$\bar{n}$	Number of individuals within each school (assumed constant across schools)
$\bar{T}$	Proportion of the sample that is assigned to the treatment group (assumed constant)
N	Total number of units $N = \sum_{k=1}^{K} \sum_{j=1}^{J} \bar{n}$
S_{id}	Categorical variable indicating the membership of individual i to a school
D_{id}	Categorical variable indicating the membership of individual i to a district
$Y_{ijmk}(0)$	Potential outcome for unit i in school j in district k for outcome m given no treatment
$Y_{ijmk}(1)$	Potential outcome for unit i in school j in district k for outcome m given treatment
V_{km}	District covariates
$g_{1,m}$	Number of district covariates for outcome m
X_{jkm}	School covariates
$g_{2,m}$	Number of school covariates for outcome m
C_{ijkm}	Individual covariates
$g_{1,m}$	Number of individual covariates for outcome m

Table 1: Observed quantities

Param	Description
$\psi_{1,jkm}$	Mean impact for school j in district k for outcome m
ξ_m	Coefficient of district covariates V_{km}
δ_m	Coefficient of school covariates X_{jkm}
γ_m	Coefficient of individual covariates C_{ijkm}

Table 2: Parameters we estimate

Param	Description
$\Xi_{0,m}$	Grand mean outcome under no treatment across districts for outcome m
$\Xi_{1,m}$	Grand mean impact across districts for outcome m
$\mu_{0,km}$	Grand mean outcome under no treatment across schools in district k for outcome m
$\mu_{1,km}$	Grand mean impact across schools in district k for outcome m
$\theta_{0,jkm}$	Mean outcome under no treatment for school j in district k for outcome m
$\psi_{1,jkm}$	Mean impact for school j in district k for outcome m
$w_{0,km}$	District intercepts
$w_{1,km}$	District impacts
$\eta^2_{0,m}$	Variance of district random effects for outcome m
$\eta^2_{1,m}$	Variance of district impacts for outcome m (cross-district treatment heterogeneity)
$u_{0,jkm}$	School intercepts
$u_{1,jkm}$	School impacts
$\tau^2_{0,m}$	Variance of school random effects for outcome m
$\tau^2_{1,m}$	Variance of school impacts for outcome m (cross-school treatment heterogeneity)
r_{ijkm}	Individual intercepts
σ^2_m	Variance of individual/level 1 residuals

Table 3: Latent parameters

Param	Description
ES_m	effect size
$\text{ICC}_{3,m}$	level 3 (district) intraclass correlation
$\omega_{3,m}$	ratio of variation of district impacts to district residuals
$\text{ICC}_{2,m}$	level 2 (school) intraclass correlation
$\omega_{2,m}$	ratio of variation of school impacts to school residuals
$R^2_{3,m}$	percent of district variation explained by level 3 (district) covariate V_{km}
$R^2_{2,m}$	percent of school variation explained by level 2 (school) covariate X_{jkm}
$R^2_{1,m}$	percent of individual variation explained by level 1 (individual covariate) C_{ijkm}

Table 4: Derived parameters

PowerUp	PUMP	Description
β_{0j}	$\theta_{0,jkm}$	Mean outcome under no treatment for school j in district k
β_{1j}	$\psi_{1,jkm}$	Mean impact for school j in district k
X_{ij}	C_{ijkm}	Individual covariates
β_{2j}	γ_m	Coefficient of individual covariates C_{ijkm}
γ_{00}	$\mu_{0,km}$	Grand mean outcome under no treatment across schools in district k
γ_{10}	$\mu_{1,km}$	Grand mean impact across schools in district k
W_{jk}	X_{jkm}	School covariates
γ_{01k}	δ_m	Coefficient of school covariates X_{jkm}
μ_{0j}	$u_{0,jkm}$	School intercepts
μ_{1j}	$u_{1,jkm}$	School impacts
$\tau^2_{2\|W}$	$\tau^2_{0,m}$	Variance of school random effects
τ^2_2	$\tau^2_{0,m} + \delta^2_m$	Overall variance of schools
$\tau^2_{T2\|W}$	$\tau^2_{1,m}$	Variance of school impacts
ρ_2	ICC_2	Intraclass correlation (unconditional) for level 2
ω_2	ω_2	Ratio of variation of impacts to residuals for level 2
τ_{2T2}	$\boldsymbol{\kappa}^w$	Correlations between school random effects and impacts
ξ_{000}	$\Xi_{0,m}$	Grand mean outcome under no treatment across districts
ξ_{100}	$\Xi_{1,m}$	Grand mean impact across districts
V_k	V_{km}	District covariates
ξ_{001}	ξ_m	Coefficient of district covariates V_{km}
ζ_{00}	$w_{0,km}$	District intercepts
ζ_{10}	$w_{1,km}$	District impacts
$\tau^2_{3\|V}$	$\eta^2_{0,m}$	Variance of district random effects
τ^2_3	$\eta^2_{0,m} + \xi^2_m$	Overall variance of districts
$\tau^2_{T3\|V}$	$\eta^2_{1,m}$	Variance of district impacts
τ_{3T3}	$\boldsymbol{\kappa}^w$	Correlations between district random effects and impacts

Table 5: Correspondence with PowerUpR!

1.4 General strategy

The same strategy is followed for all designs. First, we lay out a model for our outcomes, Y_{ijkm}. Next, we calculate the standard error of the treatment effect estimate, $\hat{\psi}_m$. When expressing the estimated treatment effect as an effect size, the standard error is given by:

$$Q_m \equiv \text{SE}\left(\hat{\text{ES}}_m\right) = \text{SE}\left(\frac{\hat{\psi}_m}{\text{VAR}}\right) = \frac{1}{VAR}\text{SE}\left(\hat{\psi}_m\right), \tag{1}$$

where VAR is some "Index Variation" that we are measuring our impacts against.

When analyzing actual data, we would, to estimate Q_m, plug in known values for $\bar{T}$, J, and $\bar{n}$. Any other parameters are replaced by sample estimates. Then, when testing the m^{th} null hypothesis, $\text{ES}_m = 0$, the test statistic for a t-test is given by

$$t_m \equiv \frac{\hat{\text{ES}}_m}{\hat{Q}_m}. \tag{2}$$

When the null is true, t_m follows a t distribution with mean 0 and degrees of freedom df_m depending on the design.

For power calculations we simply need to know, giving our assumptions on the design, a reasonable value for Q_m; with that, we can then calculate the power to detect an impact expressed in effect size units.

We next outline the R model and randomization scheme.

Explain multiplier?

A note on effect sizes. In describing the standard error of our estimators in terms of effect size, we need to carefully identify what we mean by an "effect size." We commonly think of an effect size as the size of an impact relative to some reference amount of variation. Typically we would use the variation in *overall* control group, but could also index by the within-group variation only. The former would include the student variation within each site, and also how the sites vary from each other. We argue that this definition is more natural. Unfortunately, difference in how an effect size is defined can create some tension regarding some of the power formula, as we will note in the following sections.

2 Blocked individual randomization, 2 level designs

This section of designs comprise what are usually referred to as *multisite experiments*. In a multisite experiment, we have a collection of sites (here, schools) and are able to randomize the individuals within each site into treatment and control. This allows for estimating an average impact for each site, in principle. That being said, we are usually interested in estimating some overall summary of impacts across all our sites.

Critically, there are four different estimands we might consider: the average impact for persons vs. impact for sites, and the average impact of the sample we have vs. the average impact of a population where the sample came from. When sites are equal sized, a common assumption for power calculations, the site and person average will be the same. We therefore ignore it here. For finite vs. super-population, we have to be more careful. Some estimation strategies target a finite-population estimand. In this document, the ones that do are blocked_i1_2c and blocked_i1_2f. The blocked_i1_2c estimation strategy does because it assumes a constant treatment impact; given this assumption, there is no uncertainty due to the sample itself as all samples have the same average impact by assumption. The blocked_i1_2f estimation strategy allows each school to have an individually estimated impact, but due to using fixed effects rather than random, it is evaluating the sample at hand. See [1] for further, in-depth, discussion. Those estimators that target the super-population need to take any uncertainty of the sample being representative of the super-population into account. Here, the one that does this is blocked_i1_2r, with a model of each school having an average impact drawn from some random distribution.

Regardless of the model used to analyze these data, the randomization scheme is the same. It is simple random sampling within each school, with $\bar{n}\bar{T}$ units assigned to treatment in each school. In R, we could randomize this way as so:

```
T.x <- randomizr::block_ra( S.id, prob = Tbar )
```

2.1 Blocked randomization of individuals, 2 level design, constant effects (blocked_i1_2c)

Characterization: 2-level design, fixed intercepts, constant treatment effect, no school covariates.

When we assume constant effects, each school has its own fixed intercept for the control outcome, and the treatment effect is modeled as constant across schools. We can also call this a fixed effects, constant treatment model [1]. This model allows some schools to have higher average outcomes than others (allowed for with the fixed effects), but are assuming the treatment impact is the same.

The model for estimating impacts on outcome m is given by:

$$\begin{aligned} Y_{ijkm} &= \psi_{1,m} T_{ijk} + \theta_{0,jkm} + \sum_{p=1}^{g_{1,m}} \gamma_{mp} C_{ijkmp} + r_{ijkm} \qquad (3) \\ \theta_{0,jkm} &= \mu_{0,km} + u_{0,jkm} \end{aligned}$$

with reduced form:

$$Y_{ijkm} = \psi_{1,m} T_{ijk} + \mu_{0,km} + \sum_{p=1}^{g_{1,m}} \gamma_{mp} C_{ijkmp} + u_{0,jkm} + r_{ijkm} \qquad (4)$$

and distributions:

$$r_{ijkm} \sim N\left(0, \sigma_m^2\right). \tag{5}$$

PowerUp! gives that the standard error of the treatment effect estimate *not* in effect size units is given by:

$$SE(\hat{\psi}_m) = \sqrt{\frac{1}{\bar{T}(1-\bar{T})J\bar{n}}} \cdot \sigma_m. \tag{6}$$

The degrees of freedom for our impact estimate are

$$\mathrm{df}_m = J\bar{n} - g_{1,m} - J - 1. \tag{7}$$

Converting to effect size. To convert the above to an effect size, we need to scale by overall variation. Unfortunately, under a fixed effect model, there is no natural way to express this as we have not parameterized how the individual site intercepts, the $\delta_{0,jkm}$, vary. PowerUp! therefore indexes by within group variation, which is

$$Var(Y_{ijkm}(0)|S_{id}) = \frac{\sigma_m^2}{1-R_{1,m}^2}$$

using the formula for $R_{1,m}^2$, capturing the predictive power of our individual-level covariates on the outcomes within a given school, of

$$R_{1,m}^2 = 1 - \frac{\sigma_m^2}{Var(Y_{ijkm}(0)|S_{id})}.$$

If we divide the above $SE(\hat{\psi}_m)$ formula by $\sigma_m^2/(1-R_{1,m}^2)$ we get the reported MDES formula for Q_m of

$$\tilde{Q}_m = \sqrt{\frac{1-R_{1,m}^2}{\bar{T}(1-\bar{T})J\bar{n}}},$$

with the tilde denoting that these effect size units are in terms of within-school variation, which is not often done. Equivalently, this is assuming the blocks are all homogeneous, which both goes counter to the design principles of blocking and also is known to generally not hold when evaluating schools. If we want the more classic effect size indexed by cross-site variation, we need to go further.

Assume we have an $\mathrm{ICC}_{2,m}$, an assumed measure of how much overall (control-side) variation is at the school level:

$$\mathrm{ICC}_{2,m} = 1 - \frac{Var(Y_{ijkm}(0)|S_{id}))}{Var(Y_{ijkm}(0))}.$$

This ICC is even defined for a finite sample, if we view the above as comparing the emperical (pooled) within-group variation to full variation. Rearranging this gives $Var(Y_{ijkm}(0)) = Var(Y_{ijkm}(0)|S_{id})/(1-\mathrm{ICC}_{2,m})$.

We can then plug this and the $R_{1,m}^2$ formula together to get

$$Var(Y_{ijkm}(0)) = \frac{\sigma_m^2}{1-R_{1,m}^2} \cdot \frac{1}{1-\mathrm{ICC}_{2,m}}.$$

If we use this expression to scale our SE formula, we finally obtain

$$Q_m = \sqrt{\frac{(1 - \text{ICC}_{2,m})(1 - R^2_{1,m})}{\bar{T}(1 - \bar{T})J\bar{n}}}. \tag{8}$$

Sample size formula. The sample size formulas are:

$$J = \left(\frac{M_{df}}{MDES}\right)^2 \left(\frac{(1 - \text{ICC}_{2,m})(1 - R^2_{1,m})}{\bar{n}\bar{T}(1 - \bar{T})}\right) \tag{9}$$

$$\bar{n} = \left(\frac{M_{df}}{MDES}\right)^2 \left(\frac{(1 - \text{ICC}_{2,m})(1 - R^2_{1,m})}{J\bar{T}(1 - \bar{T})}\right). \tag{10}$$

The constant effects model means that we assume no treatment variation across our sites, i.e.,:

- $\omega_{2,m} = 0$

Another way of viewing the PowerUp! formula without an ICC term is that we have

- $\text{ICC}_{2,m} = 0$

Note that code currently reflects Powerup formula, not corrected formula.

Code syntax. The R model is

```
Yobs ~ 1 + T.x + C.ijk + S.id
```

PowerUp! Differences. PowerUp assumes there is no $\text{ICC}_{2,m}$ term while we allow for it. This can be viewed as within (PowerUp!) vs. overall (this work) effect size metrics.

2.2 Blocked randomization of individuals, 2 level design, fixed effects (blocked_i1_2f)

Characterization: 2-level design, fixed intercepts, fixed slopes/treatment effects, no school covariates.

The constant effects model assumes treatment is the same for each block. If it is not, and the blocks are different sizes or have different proportions of units treated, that estimator is precision-weighted and can thus be biased. Some may instead choose to allow each school to have its own estimated impact, with a second averaging step where we calculate an overall site-average of the site specific impact estimates.

We do this by interacting our site fixed effects with treatment. Now each school has its own fixed intercept for the control outcome, and each school also has its own fixed coefficient for the treatment effect. We can also call this a fixed effects with interactions model [1].

In practice, the power calculations for this model will be the same as for constant effects, unless we allow for block size variation or variable proportion treated.

The model for estimating impacts on outcome m is given by:

$$Y_{ijkm} = \psi_{1,jkm} T_{ijk} + \theta_{0,jkm} + \sum_{p=1}^{g_{1,m}} \gamma_{mp} C_{ijkmp} + r_{ijkm} \tag{11}$$

$$\theta_{0,jkm} = \mu_{0,km} + u_{0,jkm}$$
$$\psi_{1,jkm} = \mu_{1,km} + u_{1,jkm}$$

with reduced form:

$$Y_{ijkm} = \left(\mu_{1,km} + u_{1,jkm}\right) T_{ijk} + \mu_{0,km} + \sum_{p=1}^{g_{1,m}} \gamma_{mp} C_{ijkmp} + u_{0,jkm} + r_{ijkm} \tag{12}$$

and distributions:

$$r_{ijkm} \sim N\left(0, \sigma_m^2\right). \tag{13}$$

The standard error of the treatment effect estimate (and therefore the sample size formula) are all the same as in the constant effects model:

$$Q_m = \sqrt{\frac{(1 - \mathrm{ICC}_{2,m})(1 - R^2_{1,m})}{\bar{T}(1 - \bar{T}) J \bar{n}}}. \tag{14}$$

However, the degrees of freedom are different due to the additional interaction terms we need to estimate:

$$\mathrm{df}_m = J\bar{n} - g_{1,m} - 2J. \tag{15}$$

Sample size formula. The sample size formulas are:

$$J = \left(\frac{M_{df}}{MDES}\right)^2 \left(\frac{(1 - \mathrm{ICC}_{2,m})(1 - R^2_{1,m})}{\bar{n}\bar{T}(1 - \bar{T})}\right) \tag{16}$$

$$\bar{n} = \left(\frac{M_{df}}{MDES}\right)^2 \left(\frac{(1 - \mathrm{ICC}_{2,m})(1 - R^2_{1,m})}{J\bar{T}(1 - \bar{T})}\right). \tag{17}$$

Code syntax. The R model is

```
Yobs ~ 0 + T.x:S.id - T.x + C.ijk
```

The overall treatment effect is then the average of the `T.x:S.id` interaction terms.

For now, for validation, we continue to assume $\mathrm{ICC}_{2,m} = 0$

PowerUp! Differences. Just as the constant model, PowerUp assumes there is no $\mathrm{ICC}_{2,m}$ term while we allow for it. This can be viewed as within (PowerUp!) vs. overall (this work) effect size metrics.

2.3 Blocked randomization of individuals, 2 level design, random effects (blocked_i1_2r)

Characterization: 2-level design, random intercepts, random slopes/treatment effect, school covariates for intercept. Powerup also has school covariates for slope to allow for modeling treatment effect heterogeneity.

If we are interested in generalizing from our sample to a superpopulation, we may wish to view the sample of schools themselves as representative of something larger. Then, if some schools have different average impacts than other schools, we have to account for the possibility that our sample of schools has an overall average impact different from the target population. We can account for this additional uncertainty with a random effects model that has a random effect for the school-level average impact.

The class random effects model gives each school both a random intercept for the control outcome, and a random coefficient for the treatment effect. This is also known as the RIRC model: random intercept, random coefficient.

The model for estimating impacts on outcome m is given by:

$$\begin{aligned} Y_{ijkm} &= \psi_{1,ijkm} T_{ijk} + \theta_{0,jkm} + \sum_{p=1}^{g_{1,m}} \gamma_{mp} C_{ijkmp} + r_{ijkm} \\ \theta_{0,jkm} &= \mu_{0,km} + \sum_{r=1}^{g_{2,m}} \delta_{mr} X_{jkmr} + u_{0,jkm} \\ \psi_{1,jkm} &= \mu_{1,km} + u_{1,jkm} \end{aligned} \tag{18}$$

with reduced form:

$$\begin{aligned} Y_{ijkm} = &\left(\mu_{1,km} + u_{1,jkm}\right) T_{ijk} + \mu_{0,km} \\ &+ \sum_{r=1}^{g_{2,m}} \delta_{mr} X_{jkmr} + \sum_{p=1}^{g_{1,m}} \gamma_{mp} C_{ijkmp} + u_{0,jkm} + r_{ijkm} \end{aligned} \tag{19}$$

and distributions:

$$\begin{aligned} \begin{pmatrix} u_{0,jkm} \\ u_{1,jkm} \end{pmatrix} &\sim N\left(\begin{pmatrix} \tau^2_{0,m} & \kappa^u_{mm}\tau_{0,m}\tau_{1,m} \\ \kappa^u_{mm}\tau_{1,m}\tau_{0,m} & \tau^2_{1,m} \end{pmatrix}, \begin{pmatrix} 0 \\ 0 \end{pmatrix} \right) \\ r_{ijkm} &\sim N\left(0, \sigma^2_m\right). \end{aligned} \tag{20}$$

In the fixed effects model, we set $u_{0,jkm}$ and $u_{1,jkm}$ to be fixed but constrained to have mean 0. In the random effects model, we impose Normally distributed effects with variances $\tau^2_{0,m}$ and $\tau^2_{1,m}$ and correlation κ^u_{mm}. We note that the correlation structure κ^u_{mm} does not heavily impact the distribution of the final test statistic.

We make an important note. In PowerUp!, they assume that school and district covariates also influence the treatment impact:

$$\psi_{1,jkm} = \mu_{1,km} + \sum_{r=1}^{g_{2,m}} \phi_{mr} X_{jkmr} u_{1,jkm}$$

but we do not make this assumption. The result of this is that we assume, in their notation, $R^2_{2T} = 0$, which is the percent of treatment variation explained by level 2 covariates; we are exploring none of the cross-site impact heterogeneity. This assumption affects the first term in the standard error below.

The standard error of the treatment effect estimate is given by:

$$Q_m = \sqrt{\frac{\text{ICC}_{2,m}\omega_{2,m}}{J} + \frac{(1-\text{ICC}_{2,m})(1-R^2_{1,m})}{\bar{T}(1-\bar{T})J\bar{n}}}. \tag{21}$$

Note that this formula is simply the prior formula with an additional inflation term of $\text{ICC}_{2,m}\omega_{2,m}/J$. This represents the additional uncertainty from extrapolating from our sample to the super-population. Q_m with this model, therefore, will be larger than the prior models to the extent that the schools differ in terms of their impact variation (the $\text{ICC}_{2,m}\omega_{2,m}$ term is simply the variation in the random impact terms scaled by our overall variation).

The degrees of freedom are

$$\text{df}_m = J - g_{1,m} - 1. \tag{22}$$

Sample size formula. The sample size formulas are:

$$J = \left(\frac{M_{df}}{MDES}\right)^2 \left(\text{ICC}_{2,m}\omega_{2,m} + \frac{(1-\text{ICC}_{2,m})(1-R^2_{1,m})}{\bar{T}(1-\bar{T})\bar{n}}\right) \tag{23}$$

$$\bar{n} = \frac{(1-\text{ICC}_{2,m})(1-R^2_{1,m})}{\bar{T}(1-\bar{T})\left(J\left(\frac{M_{df}}{MDES}\right)^{-2} - \text{ICC}_{2,m}\omega_{2,m}\right)} \tag{24}$$

Code syntax. The R model is

```
Yobs ~ 1 + T.x + X.jk + C.ijk + (1 + T.x | S.id)
```

PowerUp! Differences. PowerUp allows for school covariates to influence the treatment impact, while we do not allow for this. In PowerUp terms, we assume $R^2_{2T} = 0$.

3 Blocked randomization of individuals, 3 level designs

In these designs we have schools nested in districts, and students in schools. The only difference here, with the prior blocked individual randomization with 2 levels, is the third level of district. Since we are randomizing at the school level, this will only impact how we think about where variation is in terms of for effect size units.

In this context, if we are interested in the finite-sample impacts, other than for calculating our reference variation for effect sizes, the districts do not matter. We can simply use the prior 2 level designs if we lump district variation into the $\mathrm{ICC}_{2,m}$ terms.

Nesting in district does allow

3.1 Blocked randomization of individuals, 3 level design, random effects (blocked_i1_3r)

Characterization: 3-level design, random intercepts for school and district, random slopes/treatment effect for school and district, school and district covariates for intercept. Powerup also allows for school and district covariates for cross-site impact heterogeneity.

The model for estimating impacts on outcome m is given by:

$$\begin{aligned}
Y_{ijkm} &= \psi_{1,jkm} T_{ijk} + \theta_{0,jkm} + \sum_{p=1}^{g_{1,m}} \gamma_{mp} C_{ijkmp} + r_{ijkm} \\
\theta_{0,jkm} &= \mu_{0,km} + \sum_{r=1}^{g_{2,m}} \delta_{mr} X_{jkmr} + u_{0,jkm} \\
\psi_{1,jkm} &= \mu_{1,km} + u_{1,jkm} \\
\mu_{0,km} &= \Xi_{0,m} + \sum_{s=1}^{g_{3,m}} \xi_{ms} V_{kms} + w_{0,km} \\
\mu_{1,km} &= \Xi_{1,m} + w_{1,km}
\end{aligned} \tag{25}$$

with reduced form:

$$\begin{aligned}
Y_{ijkm} = {} & \left(\Xi_{1,jkm} + w_{1,km} + u_{1,jkm}\right) T_{ijk} + \Xi_{0,km} \\
& + \sum_{s=1}^{g_{3,m}} \xi_{ms} V_{kms} + \sum_{r=1}^{g_{2,m}} \delta_{mr} X_{jkmr} + \sum_{p=1}^{g_{1,m}} \gamma_{mp} C_{ijkmp} \\
& + w_{0,km} + u_{0,jkm} + r_{ijkm}
\end{aligned} \tag{26}$$

and distributions:

$$\begin{aligned}
\begin{pmatrix} u_{0,jkm} \\ u_{1,jkm} \end{pmatrix} &\sim N\left(\begin{pmatrix} \tau_{0,m}^2 & \kappa_{mm}^u \tau_{0,m} \tau_{1,m} \\ \kappa_{mm}^u \tau_{1,m} \tau_{0,m} & \tau_{1,m}^2 \end{pmatrix}, \begin{pmatrix} 0 \\ 0 \end{pmatrix} \right) \\
\begin{pmatrix} w_{0,km} \\ w_{1,km} \end{pmatrix} &\sim N\left(\begin{pmatrix} \eta_{0,m}^2 & \kappa_{mm}^w \eta_{0,m} \eta_{1,m} \\ \kappa_{mm}^w \eta_{1,m} \eta_{0,m} & \eta_{1,m}^2 \end{pmatrix}, \begin{pmatrix} 0 \\ 0 \end{pmatrix} \right) \\
r_{ijkm} &\sim N\left(0, \sigma_m^2\right).
\end{aligned} \tag{27}$$

Similar to the two-level blocked model, in PowerUp! they further assume that school and district covariates also influence the treatment impact

$$\psi_{1,jkm} = \mu_{1,km} + \sum_{r=1}^{g_{2,m}} \phi_{mr} X_{jkmr} u_{1,jkm}$$
$$\mu_{1,jkm} = \xi_{1,m} + \sum_{s=1}^{g_{3,m}} \zeta_{mr} V_{kms} w_{1,km}$$

but we do not make this assumption.

The standard error of the treatment effect estimate is given by:

$$Q_m = \sqrt{\frac{\mathrm{ICC}_{3,m}\omega_{3,m}}{K} + \frac{\mathrm{ICC}_{2,m}\omega_{2,m}}{JK} + \frac{(1-\mathrm{ICC}_{2,m}-\mathrm{ICC}_{3,m})(1-R^2_{1,m})}{\bar{T}(1-\bar{T})JK\bar{n}}}. \tag{28}$$

The degrees of freedom are

$$\mathrm{df}_m = K - g_{3,m} - 1. \tag{29}$$

Sample size formula.

$$K = \left(\frac{M_{df}}{MDES}\right)^2 \left(\mathrm{ICC}_{3,m}\omega_{3,m} + \frac{\mathrm{ICC}_{2,m}\omega_{2,m}}{J} + \frac{(1-\mathrm{ICC}_{2,m}-\mathrm{ICC}_{3,m})(1-R^2_{1,m})}{\bar{T}(1-\bar{T})J\bar{n}}\right) \tag{30}$$

$$J = \frac{(1-\mathrm{ICC}_{2,m}-\mathrm{ICC}_{3,m})(1-R^2_{1,m}) + \bar{T}(1-\bar{T})\bar{n}\mathrm{ICC}_{2,m}\omega_{2,m}+}{\bar{T}(1-\bar{T})\bar{n}\left(K\left(\frac{M_{df}}{MDES}\right)^{-2} - \mathrm{ICC}_{3,m}\omega_{3,m}\right)} \tag{31}$$

$$\bar{n} = \frac{(1-\mathrm{ICC}_{2,m}-\mathrm{ICC}_{3,m})(1-R^2_{1,m})}{\bar{T}(1-\bar{T})\left(JK\left(\frac{M_{df}}{MDES}\right)^{-2} - J\mathrm{ICC}_{3,m}\omega_{3,m} - \mathrm{ICC}_{2,m}\omega_{2,m}\right)} \tag{32}$$

Code syntax.

```
Yobs ~ 1 + T.x + V.k + X.jk + C.ijk + (1 + T.x | S.id) + (1 + T.x | D.id)
```

The randomization scheme is: simple random sampling occurs within each school, with $\bar{n}\bar{T}$ units assigned to treatment in each school.

```
T.x <- randomizr::block_ra( S.id, prob = Tbar )
```

PowerUp! Differences. PowerUp allows for school and district covariates to influence the treatment impact, while we do not allow for this. In PowerUp terms, we assume $R^2_{3T} = 0$ and $R^2_{2T} = 0$.

4 Cluster randomization, 2 level designs

4.1 Simple cluster randomization, 2 level design, random effects (simple_c2_2r)

Characterization: 2-level design, random intercepts, constant treatment effect for all schools, school covariates for intercept.

The model for estimating impacts on outcome m is given by:

$$Y_{ijkm} = \theta_{0,jkm} + \sum_{p=1}^{g_{1,m}} \gamma_{mp} C_{ijkmp} + r_{ijkm} \tag{33}$$

$$\theta_{0,jkm} = \mu_{0,km} + \psi_{1,m} T_{jk} + \sum_{r=1}^{g_{2,m}} \delta_{mr} X_{jkmr} + u_{0,jkm}$$

with reduced form:

$$Y_{ijkm} = \psi_{1,m} T_{jk} + \mu_{0,km} + \sum_{r=1}^{g_{2,m}} \delta_{mr} X_{jkmr} + \sum_{p=1}^{g_{1,m}} \gamma_{mp} C_{ijkmp} + u_{0,jkm} + r_{ijkm} \tag{34}$$

and distributions:

$$\begin{aligned} u_{0,jkm} &\sim N\left(0, \tau^2_{0,m}\right) \\ r_{ijkm} &\sim N\left(0, \sigma^2_m\right). \end{aligned} \tag{35}$$

The standard error of the treatment effect estimate is given by:

$$Q_m = \sqrt{\frac{\text{ICC}_{2,m}(1 - R^2_{2,m})}{\bar{T}(1-\bar{T})J} + \frac{(1 - \text{ICC}_{2,m})(1 - R^2_{1,m})}{\bar{T}(1-\bar{T})J\bar{n}}}. \tag{36}$$

The degrees of freedom are

$$\text{df}_m = J - g_{1,m} - 2. \tag{37}$$

The constant effects model means that we assume no treatment variation across our sites, i.e.:

- $\omega_{2,m} = 0$

Sample size formula.

$$J = \left(\frac{M_{df}}{MDES}\right)^2 \left(\frac{\text{ICC}_{2,m}(1 - R^2_{2,m})}{\bar{T}(1-\bar{T})} + \frac{(1 - \text{ICC}_{2,m})(1 - R^2_{1,m})}{\bar{T}(1-\bar{T})\bar{n}}\right) \tag{38}$$

$$\bar{n} = \frac{(1 - \text{ICC}_{2,m})(1 - R^2_{1,m})}{\bar{T}(1-\bar{T})J\left(\frac{M_{df}}{MDES}\right)^{-2} - \text{ICC}_{2,m}(1 - R^2_{2,m})} \tag{39}$$

Code syntax. The R model is

```
Yobs ~ 1 + T.x + X.jk + C.ijk + (1 | S.id)
```

The randomization scheme is: simple random sampling occurs across schools, with $J\bar{T}$ schools assigned to treatment.

```
T.x <- randomizr::cluster_ra( S.id, prob = Tbar )
```

5 Cluster randomization, 3 level designs

5.1 Simple cluster randomization, 3 level design, random effects (simple_c3_3r)

Characterization: 3-level design, random intercepts for schools and districts, constant treatment effect for all schools and districts, school and district covariates for intercept.

The model for estimating impacts on outcome m is given by:

$$\begin{aligned}
Y_{ijkm} &= \theta_{0,jkm} + \sum_{p=1}^{g_{1,m}} \gamma_{mp} C_{ijkmp} + r_{ijkm} \\
\theta_{0,jkm} &= \mu_{0,km} + \sum_{r=1}^{g_{2,m}} \delta_{mr} X_{jkmr} + u_{0,jkm} \\
\mu_{0,km} &= \Xi_{0,m} + \psi_{1,m} T_k + \sum_{s=1}^{g_{3,m}} \xi_{ms} V_{kms} + w_{0,km}
\end{aligned} \tag{40}$$

with reduced form:

$$\begin{aligned}
Y_{ijkm} = {} & \psi_{1,m} T_k + \Xi_{0,m} + \sum_{s=1}^{g_{3,m}} \xi_{ms} V_{kms} + \sum_{r=1}^{g_{2,m}} \delta_{mr} X_{jkmr} + \sum_{p=1}^{g_{1,m}} \gamma_{mp} C_{ijkmp} \\
& + w_{0,km} + u_{0,jkm} + r_{ijkm}
\end{aligned} \tag{41}$$

and distributions:

$$\begin{aligned}
u_{0,jkm} &\sim N\left(0, \tau_{0,m}^2\right) \\
w_{0,jkm} &\sim N\left(0, \eta_{0,m}^2\right) \\
r_{ijkm} &\sim N\left(0, \sigma_m^2\right).
\end{aligned} \tag{42}$$

The standard error of the treatment effect estimate is given by:

$$Q_m = \sqrt{\frac{\text{ICC}_{3,m}(1-R^2_{3,m})}{\bar{T}(1-\bar{T})K} + \frac{\text{ICC}_{2,m}(1-R^2_{2,m})}{\bar{T}(1-\bar{T})JK} + \frac{(1-\text{ICC}_{2,m}-\text{ICC}_{3,m})(1-R^2_{1,m})}{\bar{T}(1-\bar{T})JK\bar{n}}}. \tag{43}$$

The degrees of freedom are

$$\text{df}_m = K - g_{3,m} - 2. \tag{44}$$

The constant effects model means that we assume no treatment variation across our sites, i.e.:

- $\omega_{2,m} = 0$
- $\omega_{3,m} = 0$

Sample size formula.

$$K = \left(\frac{M_{df}}{MDES}\right)^2 \left(\frac{\mathrm{ICC}_{3,m}(1-R^2_{3,m})}{\bar{T}(1-\bar{T})} + \frac{\mathrm{ICC}_{2,m}(1-R^2_{2,m})}{\bar{T}(1-\bar{T})J} + \frac{(1-\mathrm{ICC}_{2,m}-\mathrm{ICC}_{3,m})(1-R^2_{1,m})}{\bar{T}(1-\bar{T})J\bar{n}}\right) \tag{45}$$

$$J = \frac{\bar{n}\mathrm{ICC}_{2,m}(1-R^2_{2,m}) + (1-\mathrm{ICC}_{2,m}-\mathrm{ICC}_{3,m})(1-R^2_{1,m})}{\bar{n}\left(\bar{T}(1-\bar{T})K\left(\frac{M_{df}}{MDES}\right)^{-2} - \mathrm{ICC}_{3,m}(1-R^2_{3,m})\right)} \tag{46}$$

$$\bar{n} = \frac{(1-\mathrm{ICC}_{2,m}-\mathrm{ICC}_{3,m})(1-R^2_{1,m})}{\bar{T}(1-\bar{T})JK\left(\frac{M_{df}}{MDES}\right)^{-2} - J\mathrm{ICC}_{3,m}(1-R^2_{3,m}) - \mathrm{ICC}_{2,m}(1-R^2_{2,m})} \tag{47}$$

Code syntax. The R model is

```
Yobs ~ 1 + T.x + V.k + X.jk + C.ijk + (1 | S.id) + (1 | D.id)
```

The randomization scheme is: simple random sampling occurs across districts, with $K\bar{T}$ districts assigned to treatment.

```
T.x <- randomizr::cluster_ra( D.id, prob = Tbar )
```

6 Blocked cluster randomization, 3 level designs

6.1 Blocked cluster randomization, fixed effects (blocked_c2_3f)

Characterization: 3-level design, random intercepts for schools, fixed intercepts for districts, fixed slopes/treatment effects for districts, constant slope/effects for schools within a district, school covariates for intercept.

The model for estimating impacts on outcome m is given by:

$$\begin{aligned}
Y_{ijkm} &= \theta_{0,jkm} + \sum_{p=1}^{g_{1,m}} \gamma_{mp} C_{ijkmp} + r_{ijkm} \\
\theta_{0,jkm} &= \mu_{0,km} + \sum_{r=1}^{g_{2,m}} \delta_{mr} X_{jkmr} + u_{0,jkm} \\
\mu_{0,km} &= \Xi_{0,m} + \psi_{1,km} T_k + w_{0,km} \\
\psi_{1,km} &= \mu_{1,km} \\
\mu_{1,km} &= \Xi_{1,m} + w_{1,km}
\end{aligned} \tag{48}$$

with reduced form:

$$\begin{aligned}
Y_{ijkm} = &\left(\Xi_{1,m} + w_{1,km}\right) T_{jk} + \Xi_{0,m} + \sum_{r=1}^{g_{2,m}} \delta_{mr} X_{jkmr} + \sum_{p=1}^{g_{1,m}} \gamma_{mp} C_{ijkmp} \\
&+ w_{0,km} + u_{0,jkm} + r_{ijkm}
\end{aligned} \tag{49}$$

and distributions:

$$\begin{aligned}
u_{0,jkm} &\sim N\left(0, \tau^2_{0,m}\right) \\
r_{ijkm} &\sim N\left(0, \sigma^2_m\right).
\end{aligned} \tag{50}$$

The standard error of the treatment effect estimate is given by:

$$Q_m = \sqrt{\frac{\text{ICC}_{2,m}(1 - R^2_{2,m})}{\bar{T}(1-\bar{T})JK} + \frac{(1 - \text{ICC}_{2,m})(1 - R^2_{1,m})}{\bar{T}(1-\bar{T})JK\bar{n}}}. \tag{51}$$

The degrees of freedom are

$$\text{df}_m = K(J-2) - g_{2,m}. \tag{52}$$

This model assumes: no variation of impacts within schools, and no variation at the district level.

- $\omega_{2,m} = 0$
- $R^2_3 = 0$
- $\text{ICC}_3 = 0$

Insert snarky comment about how probably ICC3 should not be zero here Note that code currently reflects Powerup formula, not corrected formula.

Sample size formula.

$$K = \left(\frac{M_{df}}{MDES}\right)^2 \left(\frac{\mathrm{ICC}_{2,m}(1-R^2_{2,m})}{\bar{T}(1-\bar{T})J} + \frac{(1-\mathrm{ICC}_{2,m})(1-R^2_{1,m})}{\bar{T}(1-\bar{T})J\bar{n}}\right) \tag{53}$$

$$J = \frac{\bar{n}\mathrm{ICC}_{2,m}(1-R^2_{2,m}) + (1-\mathrm{ICC}_{2,m})(1-R^2_{1,m})}{\bar{n}\bar{T}(1-\bar{T})\left(\frac{M_{df}}{MDES}\right)^{-2} K} \tag{54}$$

$$\bar{n} = \frac{(1-\mathrm{ICC}_{2,m})(1-R^2_{1,m})}{\bar{T}(1-\bar{T})JK\left(\frac{M_{df}}{MDES}\right)^{-2} - \mathrm{ICC}_{2,m}(1-R^2_{2,m})} \tag{55}$$

Code syntax. The R model is

```
Yobs ~ 0 + T.x * D.id - T.x + X.jk + C.ijk + (1 | S.id)
```

The overall treatment effect is then the average of the T.x interaction terms.

The randomization scheme is: simple random sampling occurs within each district, with $\bar{J}\bar{T}$ schools assigned to treatment in each district, where $\bar{J}$ is the average number of schools in each district.

```
T.x <- randomizr::block_and_cluster_ra( blocks = D.id, clusters = S.id, prob = Tbar )
```

6.2 Blocked cluster randomization, random effects (blocked_c2_3r)

Characterization: 3-level design, random intercepts for schools and districts, random slopes/treatment effect for districts, constant slope/effects for schools within a district, school and district covariates for intercept. Powerup also allows for district covariates for slope.

The model for estimating impacts on outcome m is given by:

$$\begin{aligned}
Y_{ijkm} &= \theta_{0,jkm} + \sum_{p=1}^{g_{1,m}} \gamma_{mp} C_{ijkmp} + r_{ijkm} \\
\theta_{0,jkm} &= \mu_{0,km} + \sum_{r=1}^{g_{2,m}} \delta_{mr} X_{jkmr} + u_{0,jkm} \\
\mu_{0,km} &= \Xi_{0,m} + \psi_{1,km} T_k + \sum_{s=1}^{g_{3,m}} \xi_{ms} V_{kms} + w_{0,km} \\
\psi_{1,jkm} &= \mu_{1,km} \\
\mu_{1,km} &= \Xi_{1,m} + w_{1,km}
\end{aligned} \tag{56}$$

with reduced form:

$$\begin{aligned}
Y_{ijkm} = {} & (\Xi_{1,m} + w_{1,km}) T_{jk} + \Xi_{0,m} \\
& + \sum_{s=1}^{g_{3,m}} \xi_{ms} V_{kms} + \sum_{r=1}^{g_{2,m}} \delta_{mr} X_{jkmr} + \sum_{p=1}^{g_{1,m}} \gamma_{mp} C_{ijkmp} \\
& + w_{0,km} + u_{0,jkm} + r_{ijkm}
\end{aligned} \tag{57}$$

and distributions:

$$\begin{aligned}
u_{0,jkm} &\sim N\left(0, \tau_{0,m}^2\right) \\
\begin{pmatrix} w_{0,km} \\ w_{1,km} \end{pmatrix} &\sim N\left(\begin{pmatrix} \eta_{0,m}^2 & \kappa_{mm}^w \eta_{0,m}\eta_{1,m} \\ \kappa_{mm}^w \eta_{1,m}\eta_{0,m} & \eta_{1,m}^2 \end{pmatrix}, \begin{pmatrix} 0 \\ 0 \end{pmatrix} \right) \\
r_{ijkm} &\sim N\left(0, \sigma_m^2\right).
\end{aligned} \tag{58}$$

Similar to other blocked models model, in PowerUp! they further assume that district covariates also influence the treatment impact

$$\mu_{1,jkm} = \xi_{1,m} + \sum_{s=1}^{g_{3,m}} \zeta_{mr} V_{kms} w_{1,km}$$

but we do not make this assumption.

The standard error of the treatment effect estimate is given by:

$$Q_m = \sqrt{\frac{\text{ICC}_{3,m}\omega_{3,m}}{K} + \frac{\text{ICC}_{2,m}(1 - R_{2,m}^2)}{\bar{T}(1-\bar{T})JK} + \frac{(1 - \text{ICC}_{2,m} - \text{ICC}_{3,m})(1 - R_{1,m}^2)}{\bar{T}(1-\bar{T})JK\bar{n}}}. \tag{59}$$

The degrees of freedom are

$$\text{df}_m = K - g_{3,m} - 1. \tag{60}$$

Parameter assumptions

- $\omega_{2,m} = 0$

Sample size formula.

$$K = \left(\frac{M_{df}}{MDES}\right)^2 \left(\mathrm{ICC}_{3,m}\omega_3 + \frac{\mathrm{ICC}_{2,m}(1-R^2_{2,m})}{\bar{T}(1-\bar{T})J} + \frac{(1-\mathrm{ICC}_{2,m}-\mathrm{ICC}_{3,m})(1-R^2_{1,m})}{\bar{T}(1-\bar{T})J\bar{n}}\right) \tag{61}$$

$$J = \frac{\bar{n}\mathrm{ICC}_{2,m}(1-R^2_{2,m}) + (1-\mathrm{ICC}_{2,m}-\mathrm{ICC}_{3,m})(1-R^2_{1,m})}{\bar{T}(1-\bar{T})\left(\bar{n}K\left(\frac{M_{df}}{MDES}\right)^{-2} - \bar{n}\mathrm{ICC}_{3,m}\omega_3\right)} \tag{62}$$

$$\bar{n} = \frac{(1-\mathrm{ICC}_{2,m}-\mathrm{ICC}_{3,m})(1-R^2_{1,m})}{\bar{T}(1-\bar{T})J\left(K\left(\frac{M_{df}}{MDES}\right)^{-2}\mathrm{ICC}_{3,m}\omega_{3,m}\right) - \mathrm{ICC}_{2,m}(1-R^2_{2,m})} \tag{63}$$

Code syntax. The R model is

```
Yobs ~ 1 + T.x + V.k + X.jk + C.ijk + (1 | S.id) + (1 + T.x | D.id)
```

The randomization scheme is: simple random sampling occurs within each district, with $\bar{J}\bar{T}$ schools assigned to treatment in each district, where $\bar{J}$ is the average number of schools in each district.

```
T.x <- randomizr::block_and_cluster_ra( blocks = D.id, clusters = S.id, prob = Tbar )
```

7 The Data Generation Process

We now discuss the assumed data generating process, indexed by parameters directly tied to the structural equations we use. Let $\boldsymbol{x}_{ijk\cdot}$ be the vector collecting all outcomes m.

The data generating process is done in the following stages, outlined below.

7.1 Determine DGP parameters

We have already discussed many parameters in Section 1.3. However, there are a few additional parameters required to generate data that do not directly feed into our equations for power or MDES, related to correlations, show in Table 7.1.

The parameters used in this section need to be picked based on desired aggregate relationships of the full data. See the next section for how to translate parameters such as ICC to the DGP parameters used below.

Param	Description
$\boldsymbol{\rho}^D$	Correlation matrix of district covariates $\boldsymbol{D}_{k\cdot}$
$\boldsymbol{\rho}^{w_0}$	Correlation matrix of district random effects $\boldsymbol{w}_{0,k\cdot}$
$\boldsymbol{\rho}^{w_1}$	Correlation matrix of district impacts $\boldsymbol{w}_{1,k\cdot}$
$\boldsymbol{\kappa}^w$	Non-symmetric matrix of correlations between district random effects and impacts, composed of entries $\{\kappa^w_{m,m'}\} = Corr(w_{0,km}, w_{1,km'})$
$\boldsymbol{\rho}^X$	Correlation matrix of school covariates $\boldsymbol{X}_{jk\cdot}$
$\boldsymbol{\rho}^{u_0}$	Correlation matrix of school random effects $\boldsymbol{u}_{0,jk\cdot}$
$\boldsymbol{\rho}^{u_1}$	Correlation matrix of school impacts $\boldsymbol{u}_{1,jk\cdot}$
$\boldsymbol{\kappa}^u$	Non-symmetric matrix of correlations between school random effects and impacts, composed of entries $\{\kappa^u_{m,m'}\} = Corr(u_{0,jkm}, u_{1,jkm'})$
$\boldsymbol{\rho}^C$	Correlation matrix of individual covariates $\boldsymbol{C}_{ijk\cdot}$
$\boldsymbol{\rho}^r$	Correlation matrix of individual residuals $\boldsymbol{r}_{ijk\cdot}$

Table 6: Correlation parameters

7.2 Generate level 3 (district) data

7.2.1 Level 3 covariates

Each outcome has its own district-level covariate, V_{km} with $k = 1, \ldots, K$ and $m = 1, \ldots, M$. We have $E(V_{km}) = 0$ and $Var(V_{km}) = 1$. We assume a correlation between outcomes, so we define $\boldsymbol{\rho}^D$ is a $M \times M$ symmetric correlation matrix, with ρ^D_{ij} is the value in row i and column j of the matrix $\boldsymbol{\rho}^D$.

$$\begin{pmatrix} D_{k1} \\ \vdots \\ V_{km} \end{pmatrix} \sim N\left[\begin{pmatrix} 0 \\ \vdots \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & \cdots & \rho^D_{1M} \\ \vdots & 1 & \vdots \\ \rho^D_{M1} & \cdots & 1 \end{pmatrix}\right].$$

7.2.2 Level 3 outcomes

Let $\mu_{0,km}$ be the grand mean outcome under no treatment for district k, and $\mu_{1,km}$ be the grand mean impact across schools for district k.

$$\mu_{0,km} = \Xi_{0,m} + \xi_m V_{km} + w_{0,km} \tag{64}$$

$$\mu_{1,km} = \Xi_{1,m} + w_{1,km} \tag{65}$$

Let $\Xi_{0,m}$ be the grand mean outcome under no treatment across all districts. Without loss of generality, we will set $\Xi_{0,m} = 0$ for all m. $\Xi_{1,km}$ is the grand mean impact across districts.

We now consider the distributions of random effects and impacts $w_{0,km}$ and $w_{1,km}$, starting with the marginal distributions. We have $E(w_{0,km}) = 0$, $Var(w_{0,km}) = \eta^2_{0,m}$, and correlation between outcomes $M \times M$ matrix $\boldsymbol{\rho}^w$.

$$\begin{pmatrix} w_{k1} \\ \vdots \\ w_{kM} \end{pmatrix} \sim N\left[\begin{pmatrix} 0 \\ \vdots \\ 0 \end{pmatrix}, \begin{pmatrix} \eta^2_{0,1} & \cdots & \rho^{w_0}_{1M}\eta_{0,1}\eta_{0,M} \\ \vdots & \ddots & \vdots \\ \rho^{w_0}_{M1}\eta_{0,M}\eta_{0,1} & \cdots & \eta^2_{0,M} \end{pmatrix}\right].$$

Similarly, we have $E(w_{1,km}) = 0$, $Var(w_{1,km}) = \eta^2_{1,m}$, and correlation between outcomes $M \times M$ matrix $\boldsymbol{\rho}^{u_1}$.

We now consider the joint distribution, $(w_{0,km}, w_{1,km})$ are bivariate normal on the margin with correlation κ^w_{mm}:

$$\begin{pmatrix} w_{0,km} \\ w_{1,km} \end{pmatrix} \sim N\left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \eta^2_{0,m} & \kappa^w_{mm}\eta_{0,m}\eta_{1,m} \\ \kappa^w_{mm}\eta_{1,m}\eta_{0,m} & \eta^2_{1,m} \end{pmatrix}\right], \tag{66}$$

But we want these values to be correlated across outcome (i.e., district average math test will be correlated with district average reading test). We therefore generate the full set of district k's random effects across all outcomes as a $2M$ vector of multivariate normal residuals:

$$(w_{0,k1}, \ldots, w_{0,kM}, w_{1,k1}, \ldots, w_{1,km}) \sim MVNorm(\vec{0}, \Sigma^w_{full})$$

with

$$\Sigma^u_{full} = \begin{pmatrix} \Sigma_{w_0} & \Sigma_w \\ \Sigma'_w & \Sigma_{w_1} \end{pmatrix}$$

and

$$\Sigma_{w_0} = \begin{pmatrix} \eta^2_{0,1} & \cdots & \rho^{w_0}_{1M}\eta_{0,1}\eta_{0,M} \\ \vdots & \ddots & \vdots \\ \rho^{w_1}_{M1}\eta_{0,M}\tau_{0,1} & \cdots & \eta^2_{0,M} \end{pmatrix}$$

$$\Sigma_{w_1} = \begin{pmatrix} \eta^2_{1,1} & \cdots & \rho^{w_1}_{1M}\eta_{1,1}\eta_{1,M} \\ \vdots & \ddots & \vdots \\ \rho^{w_1}_{M1}\eta_{1,M}\eta_{1,1} & \cdots & \eta^2_{1,M} \end{pmatrix}$$

For the form of the off-diagonal blocks, Σ_w, we first construct $M \times M$ matrix $\boldsymbol{\kappa}^w$ with entries $\{\kappa^w_{mm}\}$. The diagonals of this matrix are κ^w_{mm}, which have been previously defined as the correlation of the intercept and impact for outcome m. The off-diagonals are, for $m \neq m'$,

$$\kappa^w_{mm'} = cor(w_{0,km}, w_{1,km'}).$$

We assume these are fixed across different values of k, i.e. that the correlations are constant across districts.

For example, consider a 3×3 case:

$$\boldsymbol{\kappa}^w = \begin{pmatrix} cor(w_{0,k1}, w_{1,k1}) & cor(w_{0,k1}, w_{1,k2}) & cor(w_{0,k1}, w_{1,k3}) \\ cor(w_{0,k2}, w_{1,k1}) & cor(w_{0,k2}, w_{1,k2}) & cor(w_{0,k2}, w_{1,k3}) \\ cor(w_{0,k3}, w_{1,k1}) & cor(w_{0,k3}, w_{1,k2}) & cor(w_{0,k3}, w_{1,k3}) \end{pmatrix}$$

We note that this matrix does not necessarily have to be symmetric.

This gives

$$\Sigma_w = \begin{pmatrix} \kappa^w_{11}\eta_{0,1}\eta_{1,1} & \cdots & \kappa^w_{1M}\eta_{0,1}\eta_{1,M} \\ \vdots & \ddots & \vdots \\ \kappa^w_{M1}\eta_{0,M}\eta_{1,1} & \cdots & \kappa^w_{MM}\eta_{0,M}\eta_{1,M} \end{pmatrix}$$

Note how the diagonals correspond to the off-diagonal in Eq 66.

7.3 Generate level 2 (school) data

7.3.1 Level 2 covariates

Each outcome has its own school-level covariate. For example, school average reading and math pre-tests, used for adjusting reading and math outcomes (in practice we might imagine adjusting each outcome with both, but in the case of few clusters this might not be a good idea due to degrees of freedom issues).

Index covariates as X_{jkm} with $j = 1, \ldots, J$, $k = 1, \ldots, K$, and $m = 1, \ldots, M$. As with the district-level covariates, we have $E(X_{jkm}) = 0$ and $Var(X_{jkm}) = 1$, and $\boldsymbol{\rho}^X$ is a $M \times M$ symmetric correlation matrix.

7.3.2 Level 2 outcomes

Each school j in district k has its average outcome under no treatment $\theta_{0,jkm}$ and its average impacts $\psi_{1,jkm}$. The mean outcome and average impact for school j in district k for outcome m is

$$\theta_{0,jkm} = \mu_{0,km} + \delta_m X_{jkm} + u_{0,jkm} \tag{67}$$
$$\psi_{1,jkm} = \mu_{1,km} + u_{1,jkm} \tag{68}$$

We can easily convert from three-level to two-level models. If there are no districts, then $\mu_{0,km} = \Xi_{0,m}$ and $\mu_{1,km} = \Xi_{1,m}$ for all k. Essentially, we set $w_{km} = 0$, $w_{km} = 0$, and $\xi_m = 0$ for all k.

The $(u_{0,jkm}, u_{1,jkm})$ follow a multivariate Normal structure as in Section 7.2.2. We have $Var(u_{0,jkm}) = \tau^2_{0,m}$ and $Var(u_{1,jkm}) = \tau^2_{1,m}$. Also $Cov(\boldsymbol{u}_{0,jk\cdot}) = \boldsymbol{\rho}^{u_0}$ and $Cov(\boldsymbol{u}_{1,jk\cdot}) = \boldsymbol{\rho}^{u_1}$. Finally, they relate to each other with $Corr(u_{0,jkm}, u_{1,jkm'}) = \kappa^u_{mm'}$.

7.4 Generate level 1 (individual) data

7.4.1 Level 1 covariates

Individuals have individual level covariates, one for each outcome C_{ijkm}. For example, group-mean centered reading and math scores. We assume these are homoskedastic and have the same mean across sites. As with previous covariates, we have $E(C_{ijkm}) = 0$ and $Var(C_{ijkm}) = 1$, and $\boldsymbol{\rho}^C$ is a $M \times M$ symmetric correlation matrix.

7.4.2 Level 1 outcomes

For each outcome, the outcome model for the individual is

$$Y_{ijkm}(0) = \theta_{0,jkm} + \gamma_m C_{ijkm} + r_{ijkm} \tag{69}$$
$$Y_{ijkm}(1) = Y_{ijkm}(0) + \psi_{1,ijkm} \tag{70}$$

where $Y_{ijkm}(0)$ is potential outcome m under no treatment for individual i in school j in district k, and ψ_{ijkm} is the unit's individual causal effect.

We assume constant treatment effects for individuals in the same school, $\psi_{ijkm} = \psi_{1,jkm}$, but this assumption could be relaxed to allow for individual treatment-level heterogeneity.

As with previous covariates, we have $E(C_{ijkm}) = 0$ and $Var(C_{ijkm}) = 1$, and $\boldsymbol{\rho}^C$ is a $M \times M$ symmetric correlation matrix.

Finally, individual-level residuals are distributed $E(r_{ijkm}) = 0$ and $Var(r_{ijkm}) = 1$, and $\boldsymbol{\rho}^r$ is a $M \times M$ symmetric correlation matrix.

7.4.3 Reduced form

Putting the levels together, we have:

$$Y_{ijkm}(0) = \Xi_{0,m} + \xi_m V_{km} + \delta_m X_{jkm} + \gamma_m C_{ijkm} + w_{km} + u_{0,jkm} + r_{ijkm} \tag{71}$$
$$Y_{ijkm}(1) = Y_{ijkm}(0) + \Xi_{1,m} + z_{km} + u_{1,jkm} \tag{72}$$

7.5 Summary: Generating the full table of potential outcomes

1. For $k = 1, \ldots K$, and $m = 1, \ldots M$:
   (a) Generate district covariates V_{km}.
   (b) Generate district residuals $w_{0,km}$ and $w_{1,km}$.
   (c) Calculate grand means $\mu_{0,km}$ and $\mu_{1,km}$.
2. For $j = 1, \ldots J$, and $m = 1, \ldots M$:
   (a) Generate school covariates X_{jkm}.
   (b) Generate school residuals $u_{0,jkm}$ and $u_{1,jkm}$.
   (c) Calculate grand means $\theta_{0,jkm}$ and $\psi_{1,jkm}$.
3. For $i = 1, \ldots N$ and for $m = 1, \ldots M$:

(a) Generate individual covariates, C_{ijkm}

(b) Generate individual residuals r_{ijkm}.

(c) Generate predicted baseline outcomes ($Y_{ijkm}(0)$ without residuals).

(d) Add residuals to the predicted outcomes to get $Y_{ijkm}(0)$ and calculate $Y_{ijkm}(1)$.

7.6 Generate observed data

Once we have our full set of potential outcomes, we generate treatment assignments to generate the observed outcomes. We generate our treatment assignment, T_{ijk} for all $i = 1, \ldots, n_j$ and $j = 1, \ldots, J$ and $k = 1, \ldots, K$. Once we have our set of T_{ijk} (no matter how they were obtained) we calculate the observed outcomes

$$Y_{ijkm}^{obs} = Y_{ijkm}(0)(1 - T_{ijk}) + Y_{ijkm}(1)T_{ijk} \tag{73}$$

7.6.1 Randomization schemes

We can assign at the district, school, or individual level depending on the design we are generating data for.

- Blocked individual randomization: simple random sampling occurs within each school, with $\bar{n}\bar{T}$ units assigned to treatment in each school.
- Cluster 2-level randomization: simple random sampling occurs across schools, with $J\bar{T}$ schools assigned to treatment.
- Cluster 3-level randomization: simple random sampling occurs across districts, with $K\bar{T}$ districts assigned to treatment.
- Blocked cluster 3-level randomization: simple random sampling occurs within each district, with $J\bar{T}$ schools assigned to treatment in each district.

8 Tuning the DGP parameters

We define two main types of parameters. First, model parameters are those defined in Tables 3 and 7.1, and define the DGP. Second, control or derived parameters, defined in Table 4 indirectly tune model parameters. Control parameters are set by the user, which then influence model parameters which are fed into the DGP. The mapping of control parameters to model parameters is in Table 4.

We break our model parameters into sets:

- Set 1: $\{M, J, K, n_{jk}, \Xi_{0,m}, \boldsymbol{\rho}^D, \boldsymbol{\rho}^w, \boldsymbol{\rho}^z, \boldsymbol{\rho}^X, \boldsymbol{\rho}^u, \boldsymbol{\rho}^v, \boldsymbol{\rho}^C, \boldsymbol{\kappa}^{wz}, \boldsymbol{\kappa}^{uv}, p_j\}$ are set directly.
- Set 2: $\{N, \mu_{0,m}, \mu_{1,m}, \theta_{0,jkm}, \psi_{1,jkm}, Y_{ijkm}(0), Y_{ijkm}(1)\}$ are functions of the parameters above that are set directly.
- Set 3: $\{\Xi_{1,m}, \eta^2_{0,m}, \eta^2_{1,m}, \tau^2_{0,m}, \tau^2_{1,m}, \xi_m, \delta_m, \gamma_m\}$ are tuned through control parameters.

To translate from our control parameters to the model parameters we derive several relationships in the following.

8.1 Calculating the variation in random effects and impacts

We have variation at the individual, school, and district level. We want to be able to tune the proportion of variation in each of these levels. We are interested in the unconditional (covariate-free) ICC.

We have for the variance of the control side

$$\begin{aligned} Var_m(Y_{ijkm}(0)) &= \xi_m^2 Var_m(V_{km}) + \delta_m^2 Var_m(X_{jkm}) + \gamma_m^2 Var_m(C_{ijkm}) + \eta_{0,m}^2 + \tau_{0,m}^2 + \sigma_m^2 \\ &= \xi_m^2 + \eta_{0,m}^2 + \delta_m^2 + \tau_{0,m}^2 + \gamma_m^2 + \sigma_m^2. \end{aligned}$$

Looking at Equation 71, we see:

$$\mathrm{ICC}_{3,m} = \frac{Var(\theta_{0,km})}{Var(Y_{ijkm}(0))} = \frac{\xi_m^2 + \eta_{0,m}^2}{\xi_m^2 + \eta_{0,m}^2 + \delta_m^2 + \tau_{0,m}^2 + \gamma_m^2 + \sigma_m^2}.$$

$$\mathrm{ICC}_{2,m} = \frac{Var(\mu_{0,jkm})}{Var(Y_{ijkm}(0))} = \frac{\delta_m^2 + \tau_{0,m}^2}{\xi_m^2 + \eta_{0,m}^2 + \delta_m^2 + \tau_{0,m}^2 + \gamma_m^2 + \sigma_m^2}.$$

8.2 Calculating the covariate coefficients

8.2.1 Calculating the level 3 covariate coefficient ξ_m

The regression coefficients for the level 3 covariates, ξ_m, is dictated by the desired R^2 values. Thus, we would like to find ξ_m as a function of the level-3 R^2. We define $R^2_{3,m}$ as the proportion of variance between level 3 districts predicted by level 3 covariates.

We start with

$$\begin{aligned} R^2_{3,m} &= 1 - \frac{Var(w_{0,km})}{Var(\mu_{0,km})} \\ &= 1 - \frac{\eta_{0,m}^2}{\xi_m^2 Var(V_{km}) + \eta_{0,m}^2}, \end{aligned}$$

leading to

$$\begin{aligned} \xi_m &= \sqrt{\frac{\eta_{0,m}^2 R^2_{3,m}}{Var(V_{km})(1 - R^2_{3,m})}} \\ &= \sqrt{\frac{\eta_{0,m}^2 R^2_{3,m}}{1 - R^2_{3,m}}}. \end{aligned}$$

8.2.2 Calculating the level 2 covariate coefficient δ_m

We start with our level-2 R^2 being defined as the proportion of variance in level-2 schools explained by level-2 covariates:

$$R^2_{2,m} = 1 - \frac{Var(u_{0,jkm})}{Var(\theta_{0,jkm} \mid S_{id})}$$

where the conditioning $Var(\theta_{0,jkm} \mid S_{id})$ denotes the variance of outcomes *within* a particular district. This expands to

$$R^2_{2,m} = 1 - \frac{\tau^2_{0,m}}{\delta^2_m Var(X_{jkm} \mid S_{id}) + \tau^2_{0,m}}.$$

Since our X_{jkm} are generated independent of district, the conditional variance is the same as overall. This gives

$$R^2_{2,m} = 1 - \frac{\tau^2_{0,m}}{\delta^2_m Var(X_{jkm}) + \tau^2_{0,m}},$$

Leading to

$$\begin{aligned}
\delta_m &= \sqrt{\frac{\tau^2_{0,m} R^2_{1,m}}{Var(X_{jkm})(1 - R^2_{2,m})}} \\
&= \sqrt{\frac{\tau^2_{0,m} R^2_{2,m}}{1 - R^2_{2,m}}}.
\end{aligned}$$

8.2.3 Calculating the coefficient for the Level 1 variable (γ_m)

Similar to level 2, we start with our level 1 R^2 being defined as the proportion of level 1 variance in individuals explained by level 1 covariates:

$$R^2_{1,m} = 1 - \frac{\sigma^2_m}{var(Y_{ijkm}(0) \mid S_{id})},$$

where the conditioning denotes the variance of outcomes *within* a particular school.

We find

$$\begin{aligned}
R^2_{1,m} &= 1 - \frac{\sigma^2_m}{\gamma^2_m var(C_{ijkm} \mid S_{id}) + \sigma^2_m} \\
&= 1 - \frac{\sigma^2_m}{\gamma^2_m var(C_{ijkm}) + \sigma^2_m} \\
\gamma_m &= \sqrt{\frac{\sigma^2_m R^2_{1,m}}{var(C_{ijkm})(1 - R^2_{1,m})}} \\
&= \sqrt{\frac{R^2_{1,m}}{1 - R^2_{1,m}}}
\end{aligned}$$

8.3 Calculating the grand mean impacts $\Xi_{1,m}$

This is a function of effect size. The effect size is simply the overall impact in standard deviation units, with the standard deviation usually being the marginal standard deviation of the control side:

$$\Xi_{1,m} = ES_m \cdot SD_m(Y_{ijkm}(0))$$

where $SD_m(Y_{ijkm}(0))$ denotes the standard deviation over i, j, and k for fixed outcome m. We have already noted $Var_m(Y_{ijkm}(0)) = \xi_m^2 + \gamma_m^2 + \delta_m^2 + \eta_{0,m}^2 + \tau_{0,m}^2 + \sigma_m^2$.

8.4 Final results

We have produced a system of equations:

$$\begin{aligned}
\text{ICC}_{3,m} &= \frac{\xi_m^2 + \eta_{0,m}^2}{\xi_m^2 + \eta_{0,m}^2 + \delta_m^2 + \tau_{0,m}^2 + \gamma_m^2 + 1} \\
\text{ICC}_{2,m} &= \frac{\delta_m^2 + \tau_{0,m}^2}{\xi_m^2 + \eta_{0,m}^2 + \delta_m^2 + \tau_{0,m}^2 + \gamma_m^2 + 1} \\
\xi_m &= \sqrt{\frac{\eta_{0,m}^2 R_{3,m}^2}{1 - R_{3,m}^2}} \\
\delta_m &= \sqrt{\frac{\tau_{0,m}^2 R_{2,m}^2}{1 - R_{2,m}^2}} \\
\gamma_m &= \sqrt{\frac{R_{1,m}^2}{1 - R_{1,m}^2}}
\end{aligned}$$

We solve the system to find our model parameters:

$$\begin{aligned}
\gamma_m^2 &= \frac{R_{1,m}^2}{1 - R_{1,m}^2} \\
\delta_m^2 &= \frac{R_{2,m}^2}{1 - R_{1,m}^2} \frac{\text{ICC}_{2,m}}{1 - \text{ICC}_{3,m} - \text{ICC}_{2,m}} \\
\xi_m^2 &= \frac{R_{3,m}^2}{1 - R_{1,m}^2} \frac{\text{ICC}_{3,m}}{1 - \text{ICC}_{3,m} - \text{ICC}_{2,m}} \\
\tau_{0,m}^2 &= \frac{1 - R_{2,m}^2}{1 - R_{1,m}^2} \frac{\text{ICC}_{2,m}}{1 - \text{ICC}_{3,m} - \text{ICC}_{2,m}} \\
\eta_{0,m}^2 &= \frac{1 - R_{3,m}^2}{1 - R_{1,m}^2} \frac{\text{ICC}_{3,m}}{1 - \text{ICC}_{3,m} - \text{ICC}_{2,m}}
\end{aligned}$$

For details on the algebra, see Section 9.

And finally we set:

$$\eta_{1,m}^2 = \omega_{3,m}\left(\eta_{0,m}^2 + \xi_m^2\right)$$
$$\tau_{1,m}^2 = \omega_{2,m}\left(\tau_{0,m}^2 + \delta_m^2\right)$$

9 Appendix: Algebra

Let's start off with expressions we will later use:

$$\begin{aligned}
\tau_{0,m}^2 &= \frac{\delta_m^2(1-R_{2,m}^2)}{R_{2,m}^2} \\
\delta_m^2 + \tau_{0,m}^2 &= \delta_m^2 + \frac{\delta_m^2(1-R_{2,m}^2)}{R_{2,m}^2} \\
&= \frac{\delta_m^2 R_{2,m}^2 + \delta_m^2 - \delta_m^2 R_{2,m}^2}{R_{2,m}^2} \\
\delta_m^2 + \tau_{0,m}^2 &= \frac{\delta_m^2}{R_{2,m}^2}
\end{aligned}$$

We also note:

$$\begin{aligned}
\frac{\mathrm{ICC}_{3,m}}{\mathrm{ICC}_{2,m}} &= \frac{\xi_m^2 + \eta_{0,m}^2}{\delta_m^2 + \tau_{0,m}^2} \\
\xi_m^2 + \eta_{0,m}^2 &= \frac{\mathrm{ICC}_{3,m}(\delta_m^2 + \tau_{0,m}^2)}{\mathrm{ICC}_{2,m}} \\
&= \frac{\mathrm{ICC}_{3,m}\delta_m^2}{R_{2,m}^2\mathrm{ICC}_{2,m}}
\end{aligned}$$

And finally it's easy to re-express $\gamma_m^2 + 1$:

$$\begin{aligned}
\gamma_m &= \sqrt{\frac{R_{1,m}^2}{1-R_{1,m}^2}} \\
\gamma_m^2 + 1 &= \frac{R_{1,m}^2}{1-R_{1,m}^2} + 1 \\
&= \frac{1}{1-R_{1,m}^2}
\end{aligned}$$

Let's start by plugging some of these into our expression for ICC_2 to find δ_m:

$$
\begin{aligned}
\mathrm{ICC}_{2,m} &= \frac{\delta_m^2 + \tau_{0,m}^2}{\xi_m^2 + \eta_{0,m}^2 + \delta_m^2 + \tau_{0,m}^2 + \gamma_m^2 + 1} \\
&= \frac{\frac{\delta_m^2}{R_{2,m}^2}}{\frac{\mathrm{ICC}_{3,m}\delta_m^2}{R_{2,m}^2 \mathrm{ICC}_{2,m}} + \frac{\delta_m^2}{R_{2,m}^2} + \gamma_m^2 + 1} \\
\frac{\delta_m^2}{R_{2,m}^2} &= \mathrm{ICC}_{2,m}\left(\frac{\mathrm{ICC}_{3,m}\delta_m^2}{R_{2,m}^2 \mathrm{ICC}_{2,m}} + \frac{\delta_m^2}{R_{2,m}^2} + \gamma_m^2 + 1\right) \\
\delta_m^2 &= \mathrm{ICC}_{3,m}\delta_m^2 + \mathrm{ICC}_{2,m}\delta_m^2 + \mathrm{ICC}_{2,m} R_{2,m}^2(\gamma_m^2 + 1) \\
\delta_m^2 &= \frac{\mathrm{ICC}_{2,m} R_{2,m}^2(\gamma_m^2 + 1)}{1 - \mathrm{ICC}_{3,m} - \mathrm{ICC}_{2,m}} \\
&= \frac{\mathrm{ICC}_{2,m} R_{2,m}^2}{(1 - \mathrm{ICC}_{3,m} - \mathrm{ICC}_{2,m})(1 - R_{1,m}^2)}
\end{aligned}
$$

Proceeding by a similar method, we can use ICC_3 to find ξ_m:

$$
\xi_m^2 = \frac{\mathrm{ICC}_{3,m} R_{3,m}^2}{(1 - \mathrm{ICC}_{3,m} - \mathrm{ICC}_{2,m})(1 - R_{1,m}^2)}
$$

Now we can plug in to find $\tau_{0,m}^2$:

$$
\begin{aligned}
\tau_{0,m}^2 &= \frac{\delta_m^2(1 - R_{2,m}^2)}{R_{2,m}^2} \\
&= \frac{\mathrm{ICC}_{2,m} R_{2,m}^2}{(1 - \mathrm{ICC}_{3,m} - \mathrm{ICC}_{2,m})(1 - R_{1,m}^2)} \frac{(1 - R_{2,m}^2)}{R_{2,m}^2} \\
&= \frac{\mathrm{ICC}_{2,m}(1 - R_{2,m}^2)}{(1 - \mathrm{ICC}_{3,m} - \mathrm{ICC}_{2,m})(1 - R_{1,m}^2)}
\end{aligned}
$$

And similarly:

$$
\eta_{0,m}^2 = \frac{\mathrm{ICC}_{3,m}(1 - R_{3,m}^2)}{(1 - \mathrm{ICC}_{3,m} - \mathrm{ICC}_{2,m})(1 - R_{1,m}^2)}
$$

References

[1] L. Miratrix, M. Weiss, and B. Henderson. An applied researcher's guide to estimating effects from multisite individually randomized trials: Estimands, estimators, and estimates. *Journal of Research on Educational Effectiveness*, 2020.