PUMP Package: Technical Appendix

December 30, 2021

Contents

1 Introduction

Our package allows for power calculations across a range of common scenarios that a user might select. Each scenario is categorized by two choices. First, the user chooses the planned experimental design (e.g., clustered data, with randomization within cluster). Second, the user makes a choice of the planned analytic model they would use for the data, once they had it (e.g. a multilevel model with random impacts). To provide a concrete example, throughout we assume an education setting where we have students (level 1) nested within schools (level 2) nested, in the three-level case, within districts (level 3).

The *design* is characterized by the number of levels of nesting (e.g., students in schools, no districts, would be two levels) and the level of randomization (e.g., randomization of schools would be randomization at level two).

The *model* choices are a bit more complex, and we discuss them in detail below. In particular, for each model we support, we create a taxonomy by noting what modeling choice is used at each level of the model, and how covariates are used.

The outline of this appendix is as follows. In the introduction, we provide notation, a taxonomy for models, explain user-set parameters, and outline the power estimation strategy. The bulk of the appendix then provides detailed information about each supported scenario, including the assumed model and standard error formula. We then describe the data generating process that we used for the package validation; this section might also be useful for readers who wish to understand the models and assumptions more deeply. The final two sections provide explicit formula linking the user-specified parameters to the full set of parameters used for data generation, followed by derivations of these formula.

1.1 Model taxonomy

To define our models, we first assume a set of observed quantities, shown in Table 1, such as sample sizes, the outcomes, and covariates. For all notation, we use i to index level 1 (individuals), j to index level 2 (schools), k to index level 3 (districts), and m to index outcomes.

Param	Description
M	Number of outcomes
J	Number of level 2 units in each level 3 group (assumed constant across level 3 groups)
K	Number of level 3 units
$\bar{n}$	Number of level 1 units (assumed constant across level 2 groups)
$\bar{T}$	Proportion of the sample that is assigned to the treatment group (assumed constant across groups)
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 level 2 group
D_{id}	Categorical variable indicating the membership of individual i to a level 3 group
$Y_{ijmk}(0)$	Potential outcome for unit i in level 2 group j in level 3 group k for outcome m given no treatment
$Y_{ijmk}(1)$	Potential outcome for unit i in level 2 group j in level 3 group$k$ for outcome m given treatment
V_{km}	Level 3 covariates
$g_{3,m}$	Number of level 3 covariates for outcome m
X_{jkm}	Level 2 covariates
$g_{2,m}$	Number of level 2 covariates for outcome m
C_{ijkm}	Level 1 covariates
$g_{1,m}$	Number of level 1 covariates for outcome m

Table 1: Observed quantities

We also assume a set of unobserved, latent parameters, shown in Table 2. These parameters include intercepts, impacts, and coefficients on covariates.

Now, we can create a model taxonomy, based on modeling choices for intercepts and impacts. In particular we determine for each level:

- Whether the level 2 and level 3 intercepts are:
  - fixed ($u_{0,jkm}$ and $w_{0,jkm}$ are fixed effects)
  - random ($u_{0,jkm}$ and $w_{0,jkm}$ are considered to be Normally distributed, allowing for partial pooling)
- Whether the level 2 and level 3 treatment effects are:
  - constant, e.g. all units are modeled as having a single average impact ($u_{1,jkm} = 0$ or $w_{1,km} = 0$)
  - fixed, e.g. each unit has an individual estimated impact ($u_{1,jkm}$ are fixed effects constrained to have mean 0), with an additional mean impact.
  - random ($u_{1,jkm}$ and $w_{1,km}$ are Normally distributed around a mean impact)

In addition, for each level the user can plan to adjust for baseline covariates, unless there are fixed effects at that level or below. Note that in some models, PowerUp! includes treatment by covariate interactions, allowing for, in principle, heterogeneous treatment effects correlated with said covariates. We do not allow for this, as including treatment by covariate interactions adds

Param	Description
$\Xi_{0,m}$	Grand mean outcome under no treatment across level 3 units for outcome m
$\Xi_{1,m}$	Grand mean impact across level 3 units for outcome m
$\mu_{0,km}$	Grand mean outcome under no treatment across level 2 units in level 3 unit k for outcome m
$\mu_{1,km}$	Grand mean impact across level 2 units in level 3 unit k for outcome m
$\theta_{0,jkm}$	Mean outcome under no treatment for level 2 unit j in level 3 unit k for outcome m
$\psi_{1,jkm}$	Mean impact for level 2 unit j in level 3 unit k for outcome m
$w_{0,km}$	Level 3/District intercepts
$w_{1,km}$	Level 3/District impacts
$\eta^2_{0,m}$	Variance of level 3 random effects for outcome m
$\eta^2_{1,m}$	Variance of level 3 impacts for outcome m (cross-district treatment heterogeneity)
ξ_m	Coefficient of level 3 covariates V_{km}
$u_{0,jkm}$	Level 2/School intercepts
$u_{1,jkm}$	Level 2/School impacts
$\tau^2_{0,m}$	Variance of level 2 random effects for outcome m
$\tau^2_{1,m}$	Variance of level 2 impacts for outcome m (cross-school treatment heterogeneity)
δ_m	Coefficient of level 2 covariates X_{jkm}
r_{ijkm}	Level 1/Individual intercepts
σ^2_m	Variance of individual/level 1 residuals
γ_m	Coefficient of individual/level 1 covariates C_{ijkm}

Table 2: Latent parameters

complexity with estimation of average treatment effects, and is unlikely to help with the precision of an overall average impact estimate. We view covariate by treatment interactions as primarily for modeling treatment effect heterogeneity, which is not the goal of this package or project. When this difference between PowerUp! and our approach occurs, it is noted.

For users familiar with PowerUP, Table 3 provides a reference for translating notation between this document and PowerUpR!

1.2 Scenario naming convention

We denote the research design by d, followed by the number of levels and randomization level, so ‘d3.1‘ is a 3-level design with randomization at level 1. The model is denoted by m, followed by the level and the assumption for the intercepts, either f or r and then the assumption for the treatment impacts, c, f, or r. For example, m3ff2rc means at level 3, we assume fixed intercepts and fixed treatment impacts, and at level 2 we assume random intercepts and constant treatment impacts. The full design and model are specified by concatenating these together, e.g. d2.1_m3fc.

Examples:

- d2.1_m2rr: 2 level, individual assignment, level 2 random intercept and random treatment effect. Corresponds to PowerUP! blocked_i1_2r.
- d3.2_m3ff2rc: 3 level, level 2 assignment, level 3 fixed intercepts and fixed treatment effects, level 2 random intercepts and constant treatment effects. Corresponds to PowerUP!

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 vector for 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 vector for 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 vector for 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 3: Correspondence with PowerUpR!

blocked_c2_3f.

Table 4 shows the list of supported scenarios and their corresponding names in PowerUp!

PowerUpR!	**PUMP**
n/a	d1.1_m1c
bira2_1c	d2.1_m2fc
bira2_1f	d2.1_m2ff
bira2_1r	d2.1_m2fr
bira3_1r	d3.1_m3rr2rr
cra2_2r	d2.2_m2rc
cra3_3r	d3.3_m3rc2rc
bcra3_2f	d3.2_m3ff2rc
n/a	d3.2_m3fc2rc
bcra3_2r	d3.2_m3rr2rc

Table 4: Scenarios: designs and models

1.3 Derived parameters

To calculate power, a user must choose assumed values for some of the latent parameters. However, for certain parameters, the user may instead have more intuition about likely values of functions of these parameters, rather than the parameters themselves. For example, rather than choosing the value of the coefficient for a level 3 covariate (ξ_m), the user sets $R^2_{3,m}$, the amount of level three variation explained by covariates. These derived parameters, which are functions of unobserved parameters, are listed in Table 5.

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

Table 5: Derived parameters

We now provide further clarification on these derived parameters. For a more detailed discussion of these expressions, see Section 4.

To keep the formula for intraclass correlation coefficients (ICCs) and R^2 terms simple and clear, we assume all covariates are unit variance and group-mean centered. In particular, group-mean centering means covariates only explain variation at their level; in practice, raw lower level covariates can explain variation in higher levels, if they systematically differ by group. For using the formula, however, researchers are welcome to input total R^2 values that include the full explanatory power of all covariates for a given level.

The ICCs are unconditional Intraclass Correlations, meaning they include the variation explained by covariates. Because we assume all covariates are group-mean centered and have unit variance, we get the pairing structure of terms in the equations below.

$$\mathrm{ICC}_{3,m} = \frac{Var(\mu_{0,km})}{Var(Y_{ijkm}(0))} = \frac{\xi_m^2 + \eta_{0,m}^2}{\left(\xi_m^2 + \eta_{0,m}^2\right) + \left(\delta_m^2 + \tau_{0,m}^2\right) + \left(\gamma_m^2 + \sigma_m^2\right)} \tag{1}$$

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

The quantity ω is the ratio between the variation in average impact of a unit and the variation in the control-side mean of a unit.

$$\omega_{3,m} = \frac{Var(\mu_{1,jkm})}{Var(\mu_{0,km})} = \frac{\eta_{1,m}^2}{\xi_m^2 + \eta_{0,m}^2} \tag{3}$$

$$\omega_{2,m} = \frac{Var(\psi_{1,jkm} \mid \mu_{1,km})}{Var(\theta_{0,jkm} \mid \mu_{0,km})} = \frac{\tau_{1,m}^2}{\delta_m^2 + \tau_{0,m}^2} \tag{4}$$

The R^2 expressions are the percent of variation at a particular level predicted by covariates. The group-mean centering makes these formula only involve covariates at the same level as the R^2; this is a simplification of convenience. The standard error formula for the models in the remainder of the document are general, however, and a user would not need to group-mean center or rescale any covariate in practice.

$$R_{3,m}^2 = 1 - \frac{Var(w_{0,km})}{Var(\mu_{0,km})} = 1 - \frac{\eta_{0,m}^2}{\xi_m^2 + \eta_{0,m}^2} \tag{5}$$

$$R_{2,m}^2 = 1 - \frac{Var(u_{0,jkm})}{Var(\theta_{0,jkm} \mid D_{id})} = 1 - \frac{\tau_{0,m}^2}{\delta_m^2 + \tau_{0,m}^2} \tag{6}$$

$$R_{1,m}^2 = 1 - \frac{Var(r_{ijkm})}{Var(Y_{ijkm}(0) \mid S_{id}, D_{id})} = 1 - \frac{\sigma_m^2}{\gamma_m^2 + \sigma_m^2} \tag{7}$$

1.4 Power estimation 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 average treatment effect estimate, $\hat{\psi}_m$. When expressing the estimated treatment effect as an effect size, the standard error is given by:

$$Q_m \equiv \mathrm{SE}\left(\hat{\mathrm{ES}}_m\right) = \mathrm{SE}\left(\frac{\hat{\psi}_m}{\mathrm{VAR}}\right) = \frac{1}{VAR}\mathrm{SE}\left(\hat{\psi}_m\right), \tag{8}$$

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

$$t_m \equiv \frac{\hat{\text{ES}}_m}{\hat{Q}_m}. \tag{9}$$

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

For power calculations we calculate, given our assumptions on the design and selected model, a reasonable value for Q_m. We can then calculate the power to detect an impact expressed in effect size units.

From the power formulas, we can also calculate MDES and sample size requirements. From Dong and Maynard [2013], in general the MDES can be estimated as

$$MDES = MT_{df} \times \text{SE/VAR}$$

where MT_{df} is known as the multiplier and is the sum of two t statistics based on degrees of freedom df. For one-tailed tests, $MT_{df} = t_\alpha + t_{1-\beta}$ where α is the type I error rate and β is the desired power. For two-tailed tests, $MT_{df} = t_{\alpha/2} + t_{1-\beta}$. For more details, see Dong and Maynard [2013, page 31] or Bloom [2006, page 22]. Manipulating this expression then results in sample size formulae.

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. If the reference amount of variation is different, then the effect size, for the same absolute effect, will also be different. This concern of what the denominator is can create some tension regarding some of the power formula, as we will note in the following sections.

In particular, the effect size formula can use either the variation in *overall* control group, or just the within-group variation only. Overall variation includes the student variation within each site, but also how the sites vary from each other. Within-group variation is just this latter component. We argue that overall variation is more natural. We also believe all the formula should use the same definition of effect size.

Where this is most obviously a concern is with fixed effect regression. In particular, with overall variation, if we increase the ICC at level 2 or level 3, then there is less variation (relative to the reference variation) in level 1; thus an increased ICC will increase power for fixed effect regression. This is simply the realized the gains of a blocked experiment. If the effect size is calculated relative to within-group variation, however, this gain is not seen. We will note how this plays out explicitly in the following sections. An alternate approach is to have the R^2 measure include variance explained by the fixed effects. To make the formula more directly comparable, we do not take this route, but one can obtain the same results by selecting an appropriate R^2 and then setting the ICC to 0.

2 Scenarios

2.1 d1.1 designs: 1 level, randomization at level 1

This is the classic individually randomized experiment where we allocate some fraction of a single set of units to treatment.

The randomization scheme is simple random sampling:

```
T.x <- randomizr::simple_ra(N = nbar, prob = Tbar)
```

2.1.1 Constant effects (d1.1_m1c)

PowerUp name: Not applicable.

Design: 1-level design, randomization at level 1.

Model: constant intercepts, constant treatment effects, no school or district covariates.

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

$$Y_{ijkm} = \psi_{1,jkm} T_{ijk} + \theta_{0,jkm} + \sum_{p=1}^{g_{1,m}} \gamma_{mp} C_{ijkmp} + r_{ijkm} \tag{10}$$

The standard error of the treatment effect estimate is:

$$Q_m = \sqrt{\frac{(1 - R^2_{1,m})}{\bar{T}(1-\bar{T}) J \bar{n}}}. \tag{11}$$

The degrees of freedom are:

$$\text{df}_m = J\bar{n} - g_{1,m} - 1. \tag{12}$$

Sample size formula. The sample size formulas is:

$$\bar{n} = \left(\frac{MT_{df}}{MDES}\right)^2 \left(\frac{1 - R^2_{1,m}}{\bar{T}(1-\bar{T})}\right). \tag{13}$$

Code syntax. The R model is

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

2.2 d2.1 designs: 2 levels, randomization at level 1

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. These are also called blocked experiments, especially if the sites are viewed as fixed.

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 the 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 Miratrix et al. [2020] for further, in-depth, discussion. 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 proportion $\bar{T}$ 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 )
```

2.2.1 Constant effects (d2.1_m2fc)

PowerUp name: bira2_1c

Design: 2-level design, randomization at level 1 (blocked).

Model: 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 [Miratrix et al., 2020]. This model allows some schools to have higher average outcomes than others (allowed for with the fixed effects), but assumes the treatment impact is the same.

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

$$Y_{ijkm} = \psi_{1,m} T_{ijk} + \theta_{0,jkm} + \sum_{p=1}^{g_{1,m}} \gamma_{mp} C_{ijkmp} + r_{ijkm} \tag{14}$$

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

with reduced form:

$$Y_{ijkm} = \psi_{1,m} T_{ijk} + \mu_{0,km} + \sum_{p=1}^{g_{1,m}} \gamma_{mp} C_{ijkmp} + u_{0,jkm} + r_{ijkm} \tag{15}$$

and distributions:

$$r_{ijkm} \sim N\left(0, \sigma_m^2\right). \tag{16}$$

The standard error formula we use is

$$Q_m = \sqrt{\frac{(1 - \text{ICC}_{2,m})(1 - R^2_{1,m})}{\bar{T}(1 - \bar{T}) J \bar{n}}}. \tag{17}$$

See below for important details on this specific formula, and how it differs from PowerUp!

The degrees of freedom for our impact estimate are

$$\text{df}_m = J\bar{n} - g_{1,m} - J - 1. \tag{18}$$

Sample size formula. The sample size formulas are:

$$J = \left(\frac{MT_{df}}{MDES}\right)^2 \left(\frac{(1 - \text{ICC}_{2,m})(1 - R^2_{1,m})}{\bar{n}\bar{T}(1 - \bar{T})}\right) \tag{19}$$

$$\bar{n} = \left(\frac{MT_{df}}{MDES}\right)^2 \left(\frac{(1 - \text{ICC}_{2,m})(1 - R^2_{1,m})}{J\bar{T}(1 - \bar{T})}\right). \tag{20}$$

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

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

Difference from PowerUP. The PowerUp formula in effect assumes

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

although this can be motivated differently; see below.

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.

Remark on effect sizes. The standard error of the treatment effect estimate *not* in effect size units is (this taken from the PowerUp! documentation):

$$SE(\hat{\psi}_m) = \sqrt{\frac{1}{\bar{T}(1-\bar{T})J\bar{n}}} \cdot \sigma_m. \tag{21}$$

To convert this 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 standard error 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 $\text{ICC}_{2,m}$, an assumed measure of how much overall (control-side) variation is at the school level:

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

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

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

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

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

2.2.2 Fixed effects (d2.1_m2ff)

PowerUp name: bira2_1f

Design: 2-level design, randomization at level 1 (blocked).

Model: fixed intercepts, fixed 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, the constant effects 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 [Miratrix et al., 2020].

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:

$$\begin{aligned} Y_{ijkm} &= \psi_{1,jkm} T_{ijk} + \theta_{0,jkm} + \sum_{p=1}^{g_{1,m}} \gamma_{mp} C_{ijkmp} + r_{ijkm} \qquad (22) \\ \theta_{0,jkm} &= \mu_{0,km} + u_{0,jkm} \\ \psi_{1,jkm} &= \mu_{1,km} + u_{1,jkm} \end{aligned}$$

with reduced form:

$$Y_{ijkm} = \left(\mu_{1,km} + u_{1,jkm}\right) T_{ijk} + \mu_{0,km} + \sum_{p=1}^{g_{1,m}} \gamma_{mp} C_{ijkmp} + u_{0,jkm} + r_{ijkm} \tag{23}$$

and distributions:

$$r_{ijkm} \sim N\left(0, \sigma_m^2\right). \tag{24}$$

The standard error of the treatment effect estimate (and therefore the sample size formula) are all the same as in the constant effects model, i.e. for SE we have:

$$Q_m = \sqrt{\frac{(1 - \mathrm{ICC}_{2,m})(1 - R^2_{1,m})}{\bar{T}(1 - \bar{T}) J \bar{n}}}. \tag{25}$$

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

$$\mathrm{df}_m = J\bar{n} - g_{1,m} - 2J. \tag{26}$$

Sample size formula. The sample size formulas are:

$$J = \left(\frac{MT_{df}}{MDES}\right)^2 \left(\frac{(1 - \mathrm{ICC}_{2,m})(1 - R^2_{1,m})}{\bar{n}\bar{T}(1 - \bar{T})}\right) \tag{27}$$

$$\bar{n} = \left(\frac{MT_{df}}{MDES}\right)^2 \left(\frac{(1 - \mathrm{ICC}_{2,m})(1 - R^2_{1,m})}{J\bar{T}(1 - \bar{T})}\right). \tag{28}$$

Difference from PowerUP. Note that the PowerUp formula assumes

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

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.

PowerUp! Differences. Just as with 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.

2.2.3 Random effects (d2.1_m2fr or d2.1_m2rr)

PowerUp name: bira2_1r

Design: 2-level design, randomization at level 1 (blocked).

Model: random intercepts, random treatment effect, school covariates for intercept. PowerUp! also includes interaction terms of treatment and school covariates to allow for modeling treatment effect heterogeneity; we do not include this.

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 impacts.

The classic random effects model gives each school both a random intercept for the control average outcome (the intercept), and a random coefficient for the treatment effect. This is also known as the RIRC model: random intercept, random coefficient. Recently, researchers also use a variant of this model, the Fixed Intercept, Random Coefficient (FIRC) model to account for concerns such as varying proportions of units treated in different schools. For power calculations, they have the same performance (they are also similar in practice; see Miratrix et al. [2020]).

For RIRC, the model for estimating impacts on outcome m is given by:

$$
\begin{aligned}
Y_{ijkm} &= \psi_{1,jkm} T_{ijk} + \theta_{0,jkm} + \sum_{p=1}^{g_{1,m}} \gamma_{mp} C_{ijkmp} + r_{ijkm} && (29)\\
\theta_{0,jkm} &= \mu_{0,km} + \sum_{r=1}^{g_{2,m}} \delta_{mr} X_{jkmr} + u_{0,jkm}\\
\psi_{1,jkm} &= \mu_{1,km} + u_{1,jkm}
\end{aligned}
$$

with reduced form:

$$
\begin{aligned}
Y_{ijkm} = {} & \left(\mu_{1,km} + u_{1,jkm}\right) T_{ijk} + \mu_{0,km} && (30)\\
& + \sum_{r=1}^{g_{2,m}} \delta_{mr} X_{jkmr} + \sum_{p=1}^{g_{1,m}} \gamma_{mp} C_{ijkmp} + u_{0,jkm} + r_{ijkm}
\end{aligned}
$$

and random effect and residual distributions of:

$$
\begin{aligned}
\begin{pmatrix} u_{0,jkm} \\ u_{1,jkm} \end{pmatrix} &\sim N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \tau^2_{0,m} & \kappa^u_{mm}\tau_{0,m}\tau_{1,m} \\ \kappa^u_{mm}\tau_{1,m}\tau_{0,m} & \tau^2_{1,m} \end{pmatrix} \right) && (31)\\
r_{ijkm} &\sim N\left(0, \sigma^2_m\right).
\end{aligned}
$$

For FIRC, we only have the random effects model on the $u_{1,jkm}$, and have fixed effects for the $u_{0,jkm}$. For RIRC, we assume bivariate Normal effects with variances $\tau^2_{0,m}$ and $\tau^2_{1,m}$ and correlation κ^u_{mm}. 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, that $R^2_{2T} = 0$, where R^2_{2T} is the percent of treatment variation explained by level 2 covariates; we are exploiting none of the cross-site impact heterogeneity. This assumption affects the first term in the standard error formula below.

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

$$Q_m = \sqrt{\frac{\text{ICC}_{2,m}\omega_{2,m}}{J} + \frac{(1 - \text{ICC}_{2,m})(1 - R^2_{1,m})}{\bar{T}(1 - \bar{T})J\bar{n}}}. \tag{32}$$

Note that this formula is simply the formula for d2.1_m2fc with an additional term of $\text{ICC}_{2,m}\omega_{2,m}/J$. This term captures 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

$$\text{df}_m = J - g_{1,m} - 1. \tag{33}$$

Sample size formula. The sample size formulas are:

$$J = \left(\frac{MT_{df}}{MDES}\right)^2 \left(\text{ICC}_{2,m}\omega_{2,m} + \frac{(1 - \text{ICC}_{2,m})(1 - R^2_{1,m})}{\bar{T}(1 - \bar{T})\bar{n}}\right) \tag{34}$$

$$\bar{n} = \frac{(1 - \text{ICC}_{2,m})(1 - R^2_{1,m})}{\bar{T}(1 - \bar{T})\left(J\left(\frac{MT_{df}}{MDES}\right)^{-2} - \text{ICC}_{2,m}\omega_{2,m}\right)} \tag{35}$$

Code syntax. The R model is, for RIRC,

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

For FIRC it is

```
Yobs ~ 0 + T.x + X.jk + C.ijk + S.id + (0 + 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 $R^2_{2T} = 0$.

2.3 d2.2 designs: 2 levels, randomization at level 2

These are commonly called cluster randomized experiments, with the schools being the clusters. The randomization scheme is a simple random sample of $J\bar{T}$ schools are assigned to treatment:

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

2.3.1 Random effects (d2.2_m2rc)

PowerUp name: cra2_2r

Design: 2-level design, randomization at level 2 (clusters).

Model: random intercepts, constant treatment effect for all schools, school covariates for intercept.

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

$$Y_{ijkm} = \theta_{0,jkm} + \sum_{p=1}^{g_{1,m}} \gamma_{mp} C_{ijkmp} + r_{ijkm} \tag{36}$$

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

with reduced form:

$$Y_{ijkm} = \psi_{1,m} T_{jk} + \mu_{0,km} + \sum_{r=1}^{g_{2,m}} \delta_{mr} X_{jkmr} + \sum_{p=1}^{g_{1,m}} \gamma_{mp} C_{ijkmp} + u_{0,jkm} + r_{ijkm} \tag{37}$$

and distributions:

$$\begin{aligned} u_{0,jkm} &\sim N\left(0, \tau^2_{0,m}\right) \\ r_{ijkm} &\sim N\left(0, \sigma^2_m\right). \end{aligned} \tag{38}$$

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

$$Q_m = \sqrt{\frac{\text{ICC}_{2,m}(1 - R^2_{2,m})}{\bar{T}(1-\bar{T})J} + \frac{(1 - \text{ICC}_{2,m})(1 - R^2_{1,m})}{\bar{T}(1-\bar{T})J\bar{n}}}. \tag{39}$$

The degrees of freedom are

$$\text{df}_m = J - g_{1,m} - 2. \tag{40}$$

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

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

Sample size formula.

$$J = \left(\frac{MT_{df}}{MDES}\right)^2 \left(\frac{\bar{n}\mathrm{ICC}_{2,m}(1-R^2_{2,m}) + (1-\mathrm{ICC}_{2,m})(1-R^2_{1,m})}{\bar{T}(1-\bar{T})\bar{n}}\right) \tag{41}$$

$$\bar{n} = \frac{(1-\mathrm{ICC}_{2,m})(1-R^2_{1,m})}{\bar{T}(1-\bar{T})J\left(\frac{MT_{df}}{MDES}\right)^{-2} - \mathrm{ICC}_{2,m}(1-R^2_{2,m})} \tag{42}$$

Code syntax. The R model is

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

2.4 d3.1 designs: 3 levels, randomization at level 1

In these designs we have schools nested in districts, and students nested in schools. The only difference here, as compared to blocked individual randomization with two levels, is the third level of district. Since we are randomizing at the student level, this will only impact how we think about where variation is in terms of our 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 two level fixed effect designs if we lump district variation into the $\text{ICC}_{2,m}$ terms. In particular, one could use d2.1_m2ff or d2.1_m2fc for the three level case by just entering $\text{ICC}_{2,m} + \text{ICC}_{3,m}$ in for $\text{ICC}_{2,m}$. In fact, we cannot have district random or fixed effects given school-level fixed effects due to collinearity.

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

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

2.4.1 Random effects (d3.1_m3rr2rr)

PowerUp name: cra3_3r

Design: 3-level design, randomization at level 1 (blocked).

Model: random intercepts for district, random treatment effects for district, random intercepts for school, random effects for schools, school and district covariates for intercepts. Powerup also allows for school and district covariates for cross-site impact heterogeneity.

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

$$\begin{aligned} Y_{ijkm} &= \psi_{1,jkm} T_{ijk} + \theta_{0,jkm} + \sum_{p=1}^{g_{1,m}} \gamma_{mp} C_{ijkmp} + r_{ijkm} \\ \theta_{0,jkm} &= \mu_{0,km} + \sum_{r=1}^{g_{2,m}} \delta_{mr} X_{jkmr} + u_{0,jkm} \\ \psi_{1,jkm} &= \mu_{1,km} + u_{1,jkm} \\ \mu_{0,km} &= \Xi_{0,m} + \sum_{s=1}^{g_{3,m}} \xi_{ms} V_{kms} + w_{0,km} \\ \mu_{1,km} &= \Xi_{1,m} + w_{1,km} \end{aligned} \tag{43}$$

with reduced form:

$$\begin{aligned} Y_{ijkm} = &\left(\Xi_{1,jkm} + w_{1,km} + u_{1,jkm}\right) T_{ijk} + \Xi_{0,km} \\ &+ \sum_{s=1}^{g_{3,m}} \xi_{ms} V_{kms} + \sum_{r=1}^{g_{2,m}} \delta_{mr} X_{jkmr} + \sum_{p=1}^{g_{1,m}} \gamma_{mp} C_{ijkmp} \\ &+ w_{0,km} + u_{0,jkm} + r_{ijkm} \end{aligned} \tag{44}$$

and distributions:

$$
\begin{aligned}
\begin{pmatrix} u_{0,jkm} \\ u_{1,jkm} \end{pmatrix} &\sim N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \tau_{0,m}^2 & \kappa_{mm}^u \tau_{0,m}\tau_{1,m} \\ \kappa_{mm}^u \tau_{1,m}\tau_{0,m} & \tau_{1,m}^2 \end{pmatrix}\right) \\
\begin{pmatrix} w_{0,km} \\ w_{1,km} \end{pmatrix} &\sim N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \eta_{0,m}^2 & \kappa_{mm}^w \eta_{0,m}\eta_{1,m} \\ \kappa_{mm}^w \eta_{1,m}\eta_{0,m} & \eta_{1,m}^2 \end{pmatrix}\right) \\
r_{ijkm} &\sim N\left(0, \sigma_m^2\right).
\end{aligned}
\tag{45}
$$

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

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

but we do not make this assumption.

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

$$
Q_m = \sqrt{\frac{\text{ICC}_{3,m}\omega_{3,m}}{K} + \frac{\text{ICC}_{2,m}\omega_{2,m}}{JK} + \frac{(1-\text{ICC}_{2,m}-\text{ICC}_{3,m})(1-R_{1,m}^2)}{\bar{T}(1-\bar{T})JK\bar{n}}}. \tag{46}
$$

The degrees of freedom are

$$
\text{df}_m = K - 1. \tag{47}
$$

This is a very conservative degrees of freedom.

Sample size formula.

$$
K = \left(\frac{MT_{df}}{MDES}\right)^2 \left(\text{ICC}_{3,m}\omega_{3,m} + \frac{\text{ICC}_{2,m}\omega_{2,m}}{J} + \frac{(1-\text{ICC}_{2,m}-\text{ICC}_{3,m})(1-R_{1,m}^2)}{\bar{T}(1-\bar{T})J\bar{n}}\right) \tag{48}
$$

$$
J = \frac{(1-\text{ICC}_{2,m}-\text{ICC}_{3,m})(1-R_{1,m}^2) + \bar{T}(1-\bar{T})\bar{n}\text{ICC}_{2,m}\omega_{2,m}}{\bar{T}(1-\bar{T})\bar{n}\left(K\left(\frac{MT_{df}}{MDES}\right)^{-2} - \text{ICC}_{3,m}\omega_{3,m}\right)} \tag{49}
$$

$$
\bar{n} = \frac{(1-\text{ICC}_{2,m}-\text{ICC}_{3,m})(1-R_{1,m}^2)}{\bar{T}(1-\bar{T})\left(JK\left(\frac{MT_{df}}{MDES}\right)^{-2} - J\text{ICC}_{3,m}\omega_{3,m} - \text{ICC}_{2,m}\omega_{2,m}\right)} \tag{50}
$$

Code syntax.

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

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 $R_{3T}^2 = 0$ and $R_{2T}^2 = 0$. This also impacts our degrees of freedom formula, which is $\text{df}_m = K - 1$ instead of $\text{df}_m = K - g_{3,m} - 1$.

2.5 d3.2 designs: 3 levels, randomization at level 2

These are commonly called blocked, cluster-randomized experiments. You find these if, for example, schools are randomized within a set of districts, or teachers are randomized within a set of schools (with students as outcomes in both cases).

The randomization scheme is: simple random sampling occurs within each district, with $J\bar{T}$ schools assigned to treatment in each district. In R we have:

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

2.5.1 Fixed effects (d3.2_m3ff2rc)

PowerUp name: bcra3_2f

Design: 3-level design, randomization at level 2 (blocked cluster).

Model: fixed intercepts for districts, fixed treatment effects for districts, random intercepts for schools, constant effects for schools within a district, school covariates for intercept.

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

$$\begin{aligned} Y_{ijkm} &= \theta_{0,jkm} + \sum_{p=1}^{g_{1,m}} \gamma_{mp} C_{ijkmp} + r_{ijkm} \\ \theta_{0,jkm} &= \mu_{0,km} + \psi_{1,jkm} T_{jk} + \sum_{r=1}^{g_{2,m}} \delta_{mr} X_{jkmr} + u_{0,jkm} \\ \mu_{0,km} &= \Xi_{0,m} + w_{0,km} \\ \psi_{1,km} &= \Xi_{1,m} + w_{1,km} \end{aligned} \tag{51}$$

with reduced form:

$$\begin{aligned} Y_{ijkm} = &\left(\Xi_{1,m} + w_{1,km}\right) T_{jk} + \Xi_{0,m} + \sum_{r=1}^{g_{2,m}} \delta_{mr} X_{jkmr} + \sum_{p=1}^{g_{1,m}} \gamma_{mp} C_{ijkmp} \\ &+ w_{0,km} + u_{0,jkm} + r_{ijkm} \end{aligned} \tag{52}$$

and distributions:

$$\begin{aligned} u_{0,jkm} &\sim N\left(0, \tau^2_{0,m}\right) \\ r_{ijkm} &\sim N\left(0, \sigma^2_m\right). \end{aligned} \tag{53}$$

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

$$Q_m = \sqrt{\frac{\mathrm{ICC}_{2,m}(1 - R^2_{2,m})}{\bar{T}(1-\bar{T})JK} + \frac{(1 - \mathrm{ICC}_{2,m} - \mathrm{ICC}_{3,m})(1 - R^2_{1,m})}{\bar{T}(1-\bar{T})JK\bar{n}}}. \tag{54}$$

The degrees of freedom are

$$\mathrm{df}_m = K(J-2) - g_{2,m}. \tag{55}$$

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

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

Difference from PowerUP. Note that the PowerUp formula assumes

- $\text{ICC}_{3,m} = 0$

Sample size formula.

$$K = \left(\frac{MT_{df}}{MDES}\right)^2 \left(\frac{\text{ICC}_{2,m}(1 - R^2_{2,m})}{\bar{T}(1-\bar{T})J} + \frac{(1 - \text{ICC}_{2,m} - \text{ICC}_{3,m})(1 - R^2_{1,m})}{\bar{T}(1-\bar{T})J\bar{n}}\right) \tag{56}$$

$$J = \frac{\bar{n}\text{ICC}_{2,m}(1 - R^2_{2,m}) + (1 - \text{ICC}_{2,m} - \text{ICC}_{3,m})(1 - R^2_{1,m})}{\bar{n}\bar{T}(1-\bar{T})K\left(\frac{MT_{df}}{MDES}\right)^{-2}} \tag{57}$$

$$\bar{n} = \frac{(1 - \text{ICC}_{2,m} - \text{ICC}_{3,m})(1 - R^2_{1,m})}{\bar{T}(1-\bar{T})JK\left(\frac{MT_{df}}{MDES}\right)^{-2} - \text{ICC}_{2,m}(1 - R^2_{2,m})} \tag{58}$$

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.

2.5.2 Random effects (d3.2_m3rr2rc)

PowerUp name: bcra3_2r

Design: 3-level design, randomization at level 2 (blocked cluster).

Model: random intercepts for districts, random treatment effect for districts, random intercepts for schools, constant effects for schools within a district, school and district covariates for intercept. Powerup also allows for district covariates for treatment effects.

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

$$
\begin{aligned}
Y_{ijkm} &= \theta_{0,jkm} + \sum_{p=1}^{g_{1,m}} \gamma_{mp} C_{ijkmp} + r_{ijkm} \\
\theta_{0,jkm} &= \mu_{0,km} + \sum_{r=1}^{g_{2,m}} \delta_{mr} X_{jkmr} + u_{0,jkm} \\
\mu_{0,km} &= \Xi_{0,m} + \psi_{1,km} T_k + \sum_{s=1}^{g_{3,m}} \xi_{ms} V_{kms} + w_{0,km} \\
\psi_{1,jkm} &= \Xi_{1,m} + w_{1,km}
\end{aligned}
\tag{59}
$$

with reduced form:

$$
\begin{aligned}
Y_{ijkm} = {} & \left(\Xi_{1,m} + w_{1,km}\right) T_{jk} + \Xi_{0,m} \\
& + \sum_{s=1}^{g_{3,m}} \xi_{ms} V_{kms} + \sum_{r=1}^{g_{2,m}} \delta_{mr} X_{jkmr} + \sum_{p=1}^{g_{1,m}} \gamma_{mp} C_{ijkmp} \\
& + w_{0,km} + u_{0,jkm} + r_{ijkm}
\end{aligned}
\tag{60}
$$

and distributions:

$$
\begin{aligned}
u_{0,jkm} &\sim N\left(0, \tau^2_{0,m}\right) \\
\begin{pmatrix} w_{0,km} \\ w_{1,km} \end{pmatrix} &\sim N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \eta^2_{0,m} & \kappa^w_{mm}\eta_{0,m}\eta_{1,m} \\ \kappa^w_{mm}\eta_{1,m}\eta_{0,m} & \eta^2_{1,m} \end{pmatrix} \right) \\
r_{ijkm} &\sim N\left(0, \sigma^2_m\right).
\end{aligned}
\tag{61}
$$

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:

$$
Q_m = \sqrt{\frac{\mathrm{ICC}_{3,m}\omega_{3,m}}{K} + \frac{\mathrm{ICC}_{2,m}(1 - R^2_{2,m})}{\bar{T}(1-\bar{T})JK} + \frac{(1 - \mathrm{ICC}_{2,m} - \mathrm{ICC}_{3,m})(1 - R^2_{1,m})}{\bar{T}(1-\bar{T})JK\bar{n}}}. \tag{62}
$$

The degrees of freedom are

$$\mathrm{df}_m = K - 1. \tag{63}$$

Parameter assumptions

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

PowerUp! Differences. PowerUp allows for district covariates to influence the treatment impact, while we do not allow for this. In PowerUp terms, we assume $R^2_{3T} = 0$. This also impacts our degrees of freedom formula, which is $\mathrm{df}_m = K - 1$ instead of $\mathrm{df}_m = K - g_{3,m} - 1$.

Sample size formula.

$$K = \left(\frac{MT_{df}}{MDES}\right)^2 \left(\mathrm{ICC}_{3,m}\omega_3 + \frac{\mathrm{ICC}_{2,m}(1 - R^2_{2,m})}{\bar{T}(1-\bar{T})J} + \frac{(1 - \mathrm{ICC}_{2,m} - \mathrm{ICC}_{3,m})(1 - R^2_{1,m})}{\bar{T}(1-\bar{T})J\bar{n}}\right) \tag{64}$$

$$J = \frac{\bar{n}\mathrm{ICC}_{2,m}(1 - R^2_{2,m}) + (1 - \mathrm{ICC}_{2,m} - \mathrm{ICC}_{3,m})(1 - R^2_{1,m})}{\bar{n}\bar{T}(1-\bar{T})\left(K\left(\frac{MT_{df}}{MDES}\right)^{-2} - \mathrm{ICC}_{3,m}\omega_3\right)} \tag{65}$$

$$\bar{n} = \frac{(1 - \mathrm{ICC}_{2,m} - \mathrm{ICC}_{3,m})(1 - R^2_{1,m})}{\bar{T}(1-\bar{T})J\left(K\left(\frac{MT_{df}}{MDES}\right)^{-2} - \mathrm{ICC}_{3,m}\omega_{3,m}\right) - \mathrm{ICC}_{2,m}(1 - R^2_{2,m})} \tag{66}$$

Code syntax. The R model is

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

2.6 d3.3 designs: 3 levels, randomization at level 3

These designs have randomization at the top level. They are cluster randomized, but we can model the nesting structure within cluster.

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 )
```

2.6.1 Random effects (d3.3_m3rc2rc)

PowerUp name: cra3_3r

Design: 3-level design, randomization at level 3 (cluster).

Model: random intercepts for districts, constant treatment effects for districts, random intercepts for schools, constant treatment effects for schools, school and district covariates for intercept.

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

$$\begin{aligned} Y_{ijkm} &= \theta_{0,jkm} + \sum_{p=1}^{g_{1,m}} \gamma_{mp} C_{ijkmp} + r_{ijkm} \\ \theta_{0,jkm} &= \mu_{0,km} + \sum_{r=1}^{g_{2,m}} \delta_{mr} X_{jkmr} + u_{0,jkm} \\ \mu_{0,km} &= \Xi_{0,m} + \psi_{1,m} T_k + \sum_{s=1}^{g_{3,m}} \xi_{ms} V_{kms} + w_{0,km} \end{aligned} \tag{67}$$

with reduced form:

$$\begin{aligned} Y_{ijkm} = {} & \psi_{1,m} T_k + \Xi_{0,m} + \sum_{s=1}^{g_{3,m}} \xi_{ms} V_{kms} + \sum_{r=1}^{g_{2,m}} \delta_{mr} X_{jkmr} + \sum_{p=1}^{g_{1,m}} \gamma_{mp} C_{ijkmp} \\ & + w_{0,km} + u_{0,jkm} + r_{ijkm} \end{aligned} \tag{68}$$

and distributions:

$$\begin{aligned} u_{0,jkm} &\sim N\left(0, \tau_{0,m}^2\right) \\ w_{0,jkm} &\sim N\left(0, \eta_{0,m}^2\right) \\ r_{ijkm} &\sim N\left(0, \sigma_m^2\right). \end{aligned} \tag{69}$$

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

$$Q_m = \sqrt{\frac{\text{ICC}_{3,m}(1-R_{3,m}^2)}{\bar{T}(1-\bar{T})K} + \frac{\text{ICC}_{2,m}(1-R_{2,m}^2)}{\bar{T}(1-\bar{T})JK} + \frac{(1-\text{ICC}_{2,m}-\text{ICC}_{3,m})(1-R_{1,m}^2)}{\bar{T}(1-\bar{T})JK\bar{n}}}. \tag{70}$$

The degrees of freedom are

$$\text{df}_m = K - g_{3,m} - 2. \tag{71}$$

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.

$$K = \left(\frac{MT_{df}}{MDES}\right)^2 \left(\frac{\text{ICC}_{3,m}(1-R^2_{3,m})}{\bar{T}(1-\bar{T})} + \frac{\text{ICC}_{2,m}(1-R^2_{2,m})}{\bar{T}(1-\bar{T})J} + \frac{(1-\text{ICC}_{2,m}-\text{ICC}_{3,m})(1-R^2_{1,m})}{\bar{T}(1-\bar{T})J\bar{n}}\right) \tag{72}$$

$$J = \frac{\bar{n}\text{ICC}_{2,m}(1-R^2_{2,m}) + (1-\text{ICC}_{2,m}-\text{ICC}_{3,m})(1-R^2_{1,m})}{\bar{n}\left(\bar{T}(1-\bar{T})K\left(\frac{MT_{df}}{MDES}\right)^{-2} - \text{ICC}_{3,m}(1-R^2_{3,m})\right)} \tag{73}$$

$$\bar{n} = \frac{(1-\text{ICC}_{2,m}-\text{ICC}_{3,m})(1-R^2_{1,m})}{\bar{T}(1-\bar{T})JK\left(\frac{MT_{df}}{MDES}\right)^{-2} - J\text{ICC}_{3,m}(1-R^2_{3,m}) - \text{ICC}_{2,m}(1-R^2_{2,m})} \tag{74}$$

Code syntax. The R model is

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

3 The data generating process

We now discuss the assumed data generating process (DGP), indexed by parameters directly tied to the structural equations we use. The data generating process is done in the following stages, outlined below.

3.1 Determine DGP parameters

We have already discussed most of the required parameters in Section 1. However, there are a few additional parameters required to generate data that do not directly feed into our equations for single-outcome power or MDES, related to correlations, show in Table 3.1.

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

Param	Description
$\boldsymbol{\rho}^V$	Correlation matrix of district covariates $\boldsymbol{V}_{k\cdot}$
$\boldsymbol{\rho}^{w_0}$	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

3.2 Generate level 3 (district) data

3.2.1 Level 3 covariates

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

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

3.2.2 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.

$$\mu_{0,km} = \Xi_{0,m} + \xi_m V_{km} + w_{0,km} \tag{75}$$
$$\mu_{1,km} = \Xi_{1,m} + w_{1,km} \tag{76}$$

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 $w_{1,km}$. First, we have $E(w_{0,km}) = 0$, $Var(w_{0,km}) = \eta^2_{0,m}$, and correlation between outcomes $M \times M$ matrix $\boldsymbol{\rho}^{w_0}$.

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

Similarly, we have $E(w_{1,km}) = 0$, $Var(w_{1,km}) = \eta^2_{1,m}$, and correlation between outcomes $M \times M$ matrix $\boldsymbol{\rho}^{w_1}$.

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

$$\begin{pmatrix} w_{0,km} \\ w_{1,km} \end{pmatrix} \sim N\left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \eta^2_{0,m} & \kappa^w_{mm}\eta_{0,m}\eta_{1,m} \\ \kappa^w_{mm}\eta_{1,m}\eta_{0,m} & \eta^2_{1,m} \end{pmatrix}\right], \tag{77}$$

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 77.

3.3 Generate level 2 (school) data

3.3.1 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.

3.3.2 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

$$\theta_{0,jkm} = \mu_{0,km} + \delta_m X_{jkm} + u_{0,jkm} \quad (78)$$

$$\psi_{1,jkm} = \mu_{1,km} + u_{1,jkm} \quad (79)$$

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 3.2.2. 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 $Corr(u_{0,jkm}, u_{1,jkm'}) = \kappa^u_{mm'}$.

3.4 Generate level 1 (individual) data

3.4.1 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.

3.4.2 Level 1 outcomes

For each outcome, the outcome model for the individual is

$$Y_{ijkm}(0) = \theta_{0,jkm} + \gamma_m C_{ijkm} + r_{ijkm} \tag{80}$$
$$Y_{ijkm}(1) = Y_{ijkm}(0) + \psi_{1,ijkm} \tag{81}$$

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.

3.4.3 Reduced form

Putting the levels together, we have:

$$Y_{ijkm}(0) = \Xi_{0,m} + \xi_m V_{km} + \delta_m X_{jkm} + \gamma_m C_{ijkm} + w_{km} + u_{0,jkm} + r_{ijkm} \tag{82}$$
$$Y_{ijkm}(1) = Y_{ijkm}(0) + \Xi_{1,m} + z_{km} + u_{1,jkm} \tag{83}$$

3.5 Summary: Generating the full table of potential outcomes

1. For $k = 1, \dots K$, and $m = 1, \dots M$:
   (a) Generate district covariates V_{km}.
   (b) Generate district residuals $w_{0,km}$ and $w_{1,km}$.

(c) Calculate district grand means $\mu_{0,km}$ and $\mu_{1,km}$.

2. For $j = 1, \dots J$, and $m = 1, \dots M$:
   (a) Generate school covariates X_{jkm}.
   (b) Generate school residuals $u_{0,jkm}$ and $u_{1,jkm}$.
   (c) Calculate school grand means $\theta_{0,jkm}$ and $\psi_{1,jkm}$.
3. For $i = 1, \dots N$ and for $m = 1, \dots M$:
   (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 $Y_{ijkm}(1)$.

3.6 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 $i = 1, \dots, n_j$ and $j = 1, \dots, J$ and $k = 1, \dots, K$. Once we have our set of T_{ijk} (no matter how they were obtained) we calculate the observed outcomes

$$Y_{ijkm}^{obs} = Y_{ijkm}(0)(1 - T_{ijk}) + Y_{ijkm}(1)T_{ijk} \tag{84}$$

3.6.1 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.
- Blocked cluster 2-level randomization: school level occurs within each district, with $J\bar{T}$ schools assigned to treatment in each district.
- Cluster 3-level randomization: simple random sampling occurs across districts, with $K\bar{T}$ districts assigned to treatment.

4 Tuning the DGP parameters

We define two main types of parameters. First, model parameters are those defined in Tables 2 and 3.1, and define the DGP. Second, control or derived parameters, defined in Table 5, indirectly tune model parameters. Control parameters are set by the user, which then given the model parameters fed into the DGP. The mapping of control parameters to model parameters is in Table 5.

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 parameters 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.

4.1 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:

$$\begin{aligned} Var_m(Y_{ijkm}(0)) &= \xi_m^2 Var_m(V_{km}) + \delta_m^2 Var_m(X_{jkm}) + \gamma_m^2 Var_m(C_{ijkm}) + \eta^2_{0,m} + \tau^2_{0,m} + \sigma^2_m \\ &= \xi_m^2 + \eta^2_{0,m} + \delta_m^2 + \tau^2_{0,m} + \gamma_m^2 + \sigma^2_m. \end{aligned}$$

Looking at Equation 82, we see:

$$\text{ICC}_{3,m} = \frac{Var(\theta_{0,km})}{Var(Y_{ijkm}(0))} = \frac{\xi_m^2 + \eta^2_{0,m}}{\left(\xi_m^2 + \eta^2_{0,m}\right) + \left(\delta_m^2 + \tau^2_{0,m}\right) + \left(\gamma_m^2 + \sigma^2_m\right)}.$$

$$\text{ICC}_{2,m} = \frac{Var(\mu_{0,jkm})}{Var(Y_{ijkm}(0))} = \frac{\delta_m^2 + \tau^2_{0,m}}{\left(\xi_m^2 + \eta^2_{0,m}\right) + \left(\delta_m^2 + \tau^2_{0,m}\right) + \left(\gamma_m^2 + \sigma^2_m\right)}.$$

4.2 Calculating the covariate coefficients

4.2.1 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

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

leading to

$$\xi_m = \sqrt{\frac{\eta_{0,m}^2 R_{3,m}^2}{Var(V_{km})(1 - R_{3,m}^2)}}$$
$$= \sqrt{\frac{\eta_{0,m}^2 R_{3,m}^2}{1 - R_{3,m}^2}}.$$

4.2.2 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:

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

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

$$R_{2,m}^2 = 1 - \frac{\tau_{0,m}^2}{\delta_m^2 Var(X_{jkm} \mid S_{id}) + \tau_{0,m}^2}.$$

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

$$R_{2,m}^2 = 1 - \frac{\tau_{0,m}^2}{\delta_m^2 Var(X_{jkm}) + \tau_{0,m}^2},$$

Leading to

$$\delta_m = \sqrt{\frac{\tau_{0,m}^2 R_{1,m}^2}{Var(X_{jkm})(1 - R_{2,m}^2)}}$$
$$= \sqrt{\frac{\tau_{0,m}^2 R_{2,m}^2}{1 - R_{2,m}^2}}.$$

4.2.3 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:

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

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

We find

$$\begin{aligned}
R_{1,m}^2 &= 1 - \frac{\sigma_m^2}{\gamma_m^2 var(C_{ijkm} \mid S_{id}) + \sigma_m^2} \\
&= 1 - \frac{\sigma_m^2}{\gamma_m^2 var(C_{ijkm}) + \sigma_m^2} \\
\gamma_m &= \sqrt{\frac{\sigma_m^2 R_{1,m}^2}{var(C_{ijkm})(1 - R_{1,m}^2)}} \\
&= \sqrt{\frac{R_{1,m}^2}{1 - R_{1,m}^2}}
\end{aligned}$$

4.3 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 $Var_m(Y_{ijkm}(0)) = \xi_m^2 + \gamma_m^2 + \delta_m^2 + \eta_{0,m}^2 + \tau_{0,m}^2 + \sigma_m^2$.

4.4 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:

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

For details on the algebra, see Section 5.

And finally we set:

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

5 Appendix: Derivations of parameter formulae

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

H. S. Bloom. The core analytics of randomized experiments for social research, 2006.

N. Dong and R. Maynard. Powerup!: A tool for calculating minimum detectable effect sizes and minimum required sample sizes for experimental and quasi- experimental design studies. *Journal of Research on Educational Effectiveness*, 2013.

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.