Power Analysis for Program Evaluation
Level III: Applied Power Analysis

Power Examples Using "PowerUp!"

Background: An urban area in which the program operates is home to 10 middle schools, each with about 200 students each (2,000 students total). Since students are clustered in schools, some evaluation designs may incorporate "levels" to produce multi-level designs. In this case, the students would be considered "level 1" and the schools would be "level 2." *Available Sample*: This program is able to serve about 500 students, and wishes to evaluate its impact using a randomized trial to estimate differences between program students and comparison students in math scores. *Major Parameters*: The expected effect size of the impact is about .20 standard deviations. Extant data also show that the typical variance in student math scores between schools (level 2) is about 15 percent of the total variation (ICC = .15), and that the typical correlation between pre- and post-test scores is about $\rho_1 = .7$ for students (level 1). We also assume that school averages on the pre-test correlate with school averages on the post-test at $\rho_2 = .80$. Finally, if each school has a mix of treatment and control students, the variation in treatment impacts across schools divided by variation in school means will be a ratio of $\omega = .2$, but also that the correlation between the pre-test and the school-specific treatment effects is $\rho_T = .3$. Note that many programs ask for the R-square associated with the correlations between covariates and the outcome. This statistic is simply the square of the correlation, i.e., $R^2 = \rho^2$.

Interpretation: The following examples present the minimum detectable effect size (impact) for a variety of hypothetical designs, each using 1,000 students, 500 program, 500 control. *Each design will produce a minimum detectable effect size to be compared to the expected effect size of .2.* When the minimum detectable effect size (MDES) is *larger* than the expected effect size (MDES > .2), the study is *underpowered* because the expected effect size is not large enough relative to the sample size to produce a statistically significant finding. If the minimum detectable effect size is smaller than the expected effect size (MDES < .2), the study is *adequately powered* since the expected effect size is large enough relative to the sample size to produce a statistically significant finding.

(continued on next page)

Example #1: Simple random sample (no covariates)

Design: After recruiting 1,000 study participants, 500 students are assigned to the program and 500 to the control group. In PowerUp, this is noted as an Individual Random Assignment Design, and we wish to compute the minimum detectable effect size (MDES) for this sample and compare this to the expected effect size of .20.

Model 1.0: MDES Calculator for Individual Random Assignment (IRA) Designs—Completely Randomized Controlled Trials		
Assumptions		**Comments**
Alpha Level (α)	0.05	Probability of a Type I error
Two-tailed or One-tailed Test?	2	
Power (1-β)	0.80	Statistical power (1-probability of a Type II error)
P	0.50	Proportion of the sample randomized to treatment: n_T / (n_T + n_C)
R^2	0.00	Percent of variance in outcome explained by covariates
k*	0	Number of covariates used
n (Total Sample Size)	1000	
M (Multiplier)	2.80	Computed from T_1 and T_2
T_1 (Precision)	1.96	Determined from alpha level, given two-tailed or one-tailed test
T_2 (Power)	0.84	Determined from given power level
MDES	**0.177**	Minimum Detectable Effect Size

Sample Write-up of Results:

A power analysis was calculated to ensure that expected sample sizes adequately power the study to detect significant differences between the program students and a comparison group of students who will not participate in the program. The program typically is able to serve about 500 students each program year, so the study plans to form two evenly sized groups of 500 program students and 500 comparison students using random assignment. An Individual Random Assignment Design in PowerUp was used to compute the minimum detectable effect size associated with this sample. Based on the literature on math interventions provided to similar populations, an effect size of .20 is expected.

The power analysis (assuming an alpha of .05 and power of .80) reports that a minimum sample size of 1,000 total students, or approximately 500 students per group, is adequate to detect an effect size of at least 0.18 compared to the expected effect size of .20. Therefore, the program sample estimates confirm that the available pool of students to include in the program and comparison groups appears to be sufficient for meeting the minimum sample size requirements for properly powering the study. The study also estimates a low attrition rate of 10% based on student attrition levels reported by the program over the past few years. Efforts to include sample sizes larger than the required minimum will ensure that attrition among students does not affect the study's power.

Example #2: Simple random sample (with covariates)

Design: After recruiting 1,000 study participants, 500 students are assigned to the program and 500 are assigned to the control group. The study also collects a pre-test measure to use as a covariate with a correlation of .7 and an R-square of $.7^2$=.49. In PowerUp, this is noted as an Individual Random Assignment Design, and we wish to compute the minimum detectable effect size (MDES) for this sample and compare this to the expected effect size of .20.

Model 1.0: MDES Calculator for Individual Random Assignment (IRA) Designs—Completely Randomized Controlled Trials		
Assumptions		**Comments**
Alpha Level (α)	0.05	Probability of a Type I error
Two-tailed or One-tailed Test?	2	
Power (1-β)	0.80	Statistical power (1-probability of a Type II error)
P	0.50	Proportion of the sample randomized to treatment: n_T / (n_T + n_C)
R^2	0.49	Percent of variance in outcome explained by covariates
k*	1	Number of covariates used
n (Total Sample Size)	1000	
M (Multiplier)	2.80	Computed from T_1 and T_2
T_1 (Precision)	1.96	Determined from alpha level, given two-tailed or one-tailed test
T_2 (Power)	0.84	Determined from given power level
MDES	**0.127**	Minimum Detectable Effect Size

Sample Write-up of Results:

A power analysis was calculated to ensure that expected sample sizes adequately power the study to detect significant differences between the program students and a comparison group of students who will not participate in the program. The program typically is able to serve about 500 students each program year, so the study plans to form two evenly sized groups of 500 program students and 500 comparison students using random assignment. An Individual Random Assignment Design in PowerUp was used to compute the minimum detectable effect size associated with this sample, assuming 1 covariate based on pre-test scores with an R-square of $.7^2$=.49. Based on the literature on math interventions provided to similar populations, an effect size of .20 is expected.

The power analysis (assuming an alpha of .05 and power of .80) reports that a minimum sample size of 1,000 total students, or approximately 500 students per group, is adequate to detect an effect size of at least 0.13 compared to the expected effect size of .20. Therefore, the program sample estimates confirm that the available pool of students to include in the program and comparison groups appears to be sufficient for meeting the minimum sample size requirements for properly powering the study. The study also estimates a low attrition rate of 10% based on student attrition levels reported by the program over the past few years. Efforts to include sample sizes larger than the required minimum will ensure that attrition among students does not affect the study's power.

Example #3: Complex sample with school (level 2) assignment (no covariates)

Design: After recruiting all 20 schools and 50 study participants from each school, 10 schools are assigned to the program and 10 schools are assigned to the control group. In PowerUp, this is noted as a Two-level Cluster Random Assignment Design. We compute the minimum detectable effect size (MDES) for this sample using an ICC of .15, and compare the effect size to the expected effect size of .20.

Model 3.1: MDES Calculator for Two-Level Cluster Random Assignment Design (CRA2_2)—Treatment at Level 2		
Assumptions		**Comments**
Alpha Level (α)	0.05	Probability of a Type I error
Two-tailed or One-tailed Test?	2	
Power (1-β)	0.80	Statistical power (1-probability of a Type II error)
Rho (ICC)	0.15	Proportion of variance in outcome that is between clusters
P	0.50	Proportion of Level 2 units randomized to treatment: J_T / (J_T + J_C)
R_1^2	0.00	Proportion of variance in Level 1 outcomes explained by Level 1 covariates
R_2^2	0.00	Proportion of variance in Level 2 outcome explained by Level 2 covariates
g*	0	Number of Level 2 covariates
n (Average Cluster Size)	50	Mean number of Level 1 units per Level 2 cluster (harmonic mean recommended)
J (Sample Size [# of Clusters])	20	Number of Level 2 units
M (Multiplier)	2.96	Computed from T_1 and T_2
T_1 (Precision)	2.10	Determined from alpha level, given two-tailed or one-tailed test
T_2 (Power)	0.86	Determined from given power level
MDES	**0.542**	Minimum Detectable Effect Size

Sample Write-up of Results:

A power analysis was calculated to ensure that expected sample sizes adequately power the study to detect significant differences between the program students and a comparison group of students who will not participate in the program. The program has the capacity to serve about 50 students at each of 10 schools (500 students total) and recruited 20 schools to participate in the study. The study will randomly determine which schools receive the program by forming two evenly sized groups of 10 program schools and 10 comparison schools using random assignment. A Two-level Cluster Random Assignment Design in PowerUp was used to compute the minimum detectable effect size associated with this sample, assuming an ICC of .15. Based on the literature on math interventions provided to similar populations, an effect size of .20 is expected.

The power analysis (assuming an alpha of .05 and power of .80) reports that a minimum sample size of 50 students at each of 20 schools (1,000 students total), or approximately 500 students across 10 schools per group, can detect an effect size of at least 0.54 compared to the expected smaller effect size of .20. Therefore, the program sample estimates indicate that a much larger pool of schools and students are required for meeting the minimum sample size requirements for properly powering the study. The study will explore other design options and/or potential methods for increasing the sample size in order to reduce the MDES.

Example #4: Complex sample with school (level 2) assignment (with covariates)

Design: After recruiting all 20 schools and 50 study participants from each school, 10 schools are assigned to the program and 10 schools are assigned to the control group. In PowerUp, this is noted as a Two-level Cluster Random Assignment Design. The study also collects a pre-test measure to use as a covariate, which is typically correlated with the post-test at .7. We compute the minimum detectable effect size (MDES) for this sample using an ICC of .15, and compare the effect size to the expected effect size of .20. We use a covariate at the student level with an R-square of $.7^2$=.49 and another covariate at the school level, making the additional assumption that school averages on the pre-test correlate with school averages on the post-test at .8, for an R-square of $.8^2$=.64.

Model 3.1: MDES Calculator for Two-Level Cluster Random Assignment Design (CRA2_2)—Treatment at Level 2		
Assumptions		**Comments**
Alpha Level (α)	0.05	Probability of a Type I error
Two-tailed or One-tailed Test?	2	
Power (1-β)	0.80	Statistical power (1-probability of a Type II error)
Rho (ICC)	0.15	Proportion of variance in outcome that is between clusters
P	0.50	Proportion of Level 2 units randomized to treatment: J_T / (J_T + J_C)
R_1^2	0.49	Proportion of variance in Level 1 outcomes explained by Level 1 covariates
R_2^2	0.64	Proportion of variance in Level 2 outcome explained by Level 2 covariates
g*	1	Number of Level 2 covariates
n (Average Cluster Size)	50	Mean number of Level 1 units per Level 2 cluster (harmonic mean recommended)
J (Sample Size [# of Clusters])	20	Number of Level 2 units
M (Multiplier)	2.97	Computed from T_1 and T_2
T_1 (Precision)	2.11	Determined from alpha level, given two-tailed or one-tailed test
T_2 (Power)	0.86	Determined from given power level
MDES	**0.333**	Minimum Detectable Effect Size

Sample Write-up of Results:

A power analysis was calculated to ensure that expected sample sizes adequately power the study to detect significant differences between the program students and a comparison group of students who will not participate in the program. The program has the capacity to serve about 50 students at each of 10 schools (500 students total) and recruited 20 schools to participate in the study. The study will randomly determine which schools receive the program by forming two evenly sized groups of 10 program schools and 10 comparison schools using random assignment. A Two-level Cluster Random Assignment Design in PowerUp was used to compute the minimum detectable effect size associated with this sample, assuming an ICC of .15, 1 covariate with an R-square of $.7^2$=.49 at the student level, and a school level covariate with an R-square of $.8^2$=.64. Based on the literature on math interventions provided to similar populations, an effect size of .20 is expected.

The power analysis (assuming an alpha of .05 and power of .80) reports that a sample size of 50 students at each of 20 schools (1,000 students total), or approximately 500 students across 10 schools per group, can detect an effect size of at least 0.33 compared to the smaller expected effect size of .20. Therefore, the power analysis estimates indicate that a much larger pool of schools and students are required for meeting the minimum sample size requirements for properly powering the study. The study will explore other design options and/or potential methods for increasing the sample size in order to reduce the MDES.

Example #5: Complex sample with student (level 1) assignment (random effects, no covariates)

Design: After recruiting all 20 schools and 50 study participants from each school, 25 students in each school are assigned to the program and 25 students in each school are assigned to the control group. In PowerUp, this is noted as a Two-level Random Effects Blocked Individual Random Assignment Design, and we wish to compute the minimum detectable effect size (MDES) for this sample, using an ICC of .15, and compare this to the expected effect size of .20. Also, since the treatment impact will randomly vary in our analysis, we assume the variation in treatment impacts across schools divided by variation in school means would be a ratio of $\omega = .2$.

Model 2.3: MDES Calculator for 2-Level Random Effects Blocked Individual Random Assignment (BIRA2_1r) Designs— Individuals Randomized within Blocks		
Assumptions		**Comments**
Alpha Level (α)	0.05	Probability of a Type I error
Two-tailed or One-tailed Test?	2	
Power (1-β)	0.80	Statistical power (1-probability of a Type II error)
Rho (ICC)	0.15	Proportion of variance in outcome between clusters
ω	0.20	Treatment effect heterogeneity: variability in treatment effects across Level 2 units, standardized by the variability in the Level-2 outcome
P	0.50	Proportion of Level 1 units randomized to treatment: $n_T / (n_T + n_C)$
R_1^2	0.00	Proportion of variance in the Level 1 outcome explained by Level 1 covariates
R_{2T}^2	0.00	Proportion of between block variance in treatment effect explained by Level 2 covariates
g*	0	Number of Level 2 covariates
n (Average Block Size)	50	Mean number of Level 1 units per Level 2 cluster (harmonic mean recommended)
J (Sample Size [# of Blocks])	20	Number of Level 2 units in the sample
M (Multiplier)	2.95	Computed from T_1 and T_2
T_1 (Precision)	2.09	Determined from alpha level, given two-tailed or one-tailed test
T_2 (Power)	0.86	Determined from given power level
MDES	**0.207**	Minimum Detectable Effect Size

Sample Write-up of Results:

A power analysis was calculated to ensure that expected sample sizes adequately power the study to detect significant differences between the program students and a comparison group of students who will not participate in the program. The program typically is able to serve about 500 students each program year, so the study plans to form two evenly sized groups of 500 program students and 500 comparison students using random assignment. Within each of the 20 schools participating in the study, 25 students will be assigned to the program and 25 students to the control group. A Two-level Random Effects Blocked Individual Random Assignment Design in PowerUp was used to compute the minimum detectable effect size (MDES) associated with this sample, assuming an ICC of .15. Based on the literature on math interventions provided to similar populations, an effect size of .20.

The power analysis (assuming an alpha of .05 and power of .80) reports that a sample size of 1,000 total students, or approximately 500 students per group, can detect an effect size of at least 0.21 compared to the expected effect size of .20. Therefore, the program sample estimates indicate that a slightly larger pool of schools and/or students are required for meeting the minimum sample size requirements for properly powering the study. The study will explore other design options and/or potential methods for slightly increasing the sample size in order to reduce the MDES.

Example #6: Complex sample with student (level 1) assignment (random effects with covariates)

Design: After recruiting all 20 schools, and 50 study participants from each school, 25 students in each school are assigned to the program and 25 students in each school are assigned as controls. In PowerUp, this is noted as a Two-level Random Effects Blocked Individual Random Assignment Design, and we wish to compute the minimum detectable effect size (MDES) for this sample, using an ICC of .15, and compare this to the expected effect size of .20. Also, since the treatment impact will randomly vary in our analysis, we assume the variation in treatment impacts across schools divided by variation in school means would be a ratio of $\omega = .2$. We use 1 covariate at the student level with an R-square of $.7^2$=.49 and make the additional assumption that school averages on the pre-test correlate with school treatment impacts on the post-test at .3, for an R-square on the treatment effect of $.3^2$=.09.

Model 2.3: MDES Calculator for 2-Level Random Effects Blocked Individual Random Assignment (BIRA2_1r) Designs— Individuals Randomized within Blocks		
Assumptions		**Comments**
Alpha Level (α)	0.05	Probability of a Type I error
Two-tailed or One-tailed Test?	2	
Power (1-β)	0.80	Statistical power (1-probability of a Type II error)
Rho (ICC)	0.15	Proportion of variance in outcome between clusters
ω	0.20	Treatment effect heterogeneity: variability in treatment effects across Level 2 units, standardized by the variability in the Level-2 outcome
P	0.50	Proportion of Level 1 units randomized to treatment: $n_T / (n_T + n_C)$
R_1^2	0.49	Proportion of variance in the Level 1 outcome explained by Level 1 covariates
R_{2T}^2	0.09	Proportion of between block variance in treatment effect explained by Level 2 covariates
g*	1	Number of Level 2 covariates
n (Average Block Size)	50	Mean number of Level 1 units per Level 2 cluster (harmonic mean recommended)
J (Sample Size [# of Blocks])	20	Number of Level 2 units in the sample
M (Multiplier)	2.96	Computed from T_1 and T_2
T_1 (Precision)	2.10	Determined from alpha level, given two-tailed or one-tailed test
T_2 (Power)	0.86	Determined from given power level
MDES	**0.165**	Minimum Detectable Effect Size

Sample Write-up of Results:

A power analysis was calculated to ensure that expected sample sizes adequately power the study to detect significant differences between the program students and a comparison group of students who will not participate in the program. The program typically is able to serve about 500 students each program year, so the study plans to form two evenly sized groups of 500 program students and 500 comparison students using random assignment. Within each of the 20 schools participating in the study, 25 students will be assigned to the program and 25 students to the control group. A Two-level Random

Effects Blocked Individual Random Assignment Design in PowerUp was used to compute the minimum detectable effect size (MDES) associated with this sample, assuming an ICC of .15 and 1 covariate with an R-square of $.7^2$=.49. We made the additional assumption that the correlation between school averages on the pre-test and treatment effects is .3 for an R-square of $.3^2$=.09. Based on the literature on math interventions provided to similar populations, an effect size of .20 is expected.

The power analysis (assuming an alpha of .05 and power at .80) reports that a minimum sample size of 1,000 total students, or approximately 500 students per group, is adequate to detect an effect size of at least 0.17 compared to the expected effect size of .20. Therefore, the program sample estimates confirm that the available pool of students to include in the program and comparison groups appears to be sufficient for meeting the minimum sample size requirements for properly powering the study. The study also estimates a low attrition rate of 10% based on student attrition levels reported by the program over the past few years. Efforts to include sample sizes larger than the required minimum will ensure that attrition among students does not affect the study's power.

Example #7: Complex sample with student (level 1) assignment (fixed effects, no covariates)

Design: After recruiting all 20 schools and 50 study participants from each school, 25 students in each school are assigned to the program and 25 students in each school are assigned to the control group. The analysis will assume the same (fixed) treatment effect for all schools rather than a random effect. In PowerUp, this is noted as a Two-level Fixed Effects Blocked Individual Random Assignment Design, and we wish to compute the minimum detectable effect size with this sample, and compare this to the expected effect size of .20.

Note: Using a fixed effect will increase power relative to random effects, but this limits the generalizability of the findings to just the study participants. Random effects, which have lower power, allow generalizations beyond study participants.

Model 2.2: MDES Calculator for 2-Level Fixed Effects Blocked Individual Random Assignment Designs (BIRA2_1f)— School Intercepts and Interactions with TREATMENT		
Assumptions		**Comments**
Alpha Level (α)	0.05	Probability of a Type I error
Two-tailed or One-tailed Test?	2	
Power (1-β)	0.80	Statistical power (1-probability of a Type II error)
P	0.50	Proportion of Level 1 units randomized to treatment: n_T / (n_T + n_C)
R_1^2	0.00	Proportion of variance in Level 1 outcome explained by Block and Level 1 covariates
g*	0	Number of Level 1 covariates
n (Average Block Size)	50	Mean number of Level 1 units per Level 2 cluster (harmonic mean recommended)
J (Sample Size [# of Blocks])	20	Number of Level 2 units
M (Multiplier)	2.80	Computed from T_1 and T_2
T_1 (Precision)	1.96	Determined from alpha level, given two-tailed or one-tailed test
T_2 (Power)	0.84	Determined from given power level
MDES	**0.177**	Minimum Detectable Effect Size

Sample Write-up of Results:

A power analysis was calculated to ensure that expected sample sizes adequately power the study to detect significant differences between the program students and a comparison group of students who will not participate in the program. The program typically is able to serve about 500 students each program year, so the study plans to form two evenly sized groups of 500 program students and 500 comparison students using random assignment. Within each of the 20 schools participating in the study, 25 students will be assigned to the program and 25 students to the control group. A Two-level Fixed Effects Blocked Individual Random Assignment Design in PowerUp was used to compute the minimum detectable effect size (MDES) associated with this sample. Based on the literature on math interventions provided to similar populations, an effect size of .20 is assumed.

The power analysis (assuming an alpha of .05 and power of .80) reports that a sample size of 1,000 total students, or approximately 500 students per group, is adequate to detect an effect size of at least 0.18

compared to the expected effect size of .20. Therefore, the program sample estimates confirm that the available pool of students to include in the program and comparison groups appears to be sufficient for meeting the minimum sample size requirements for properly powering the study. The study also estimates a low attrition rate of 10% based on student attrition levels reported by the program over the past few years. Efforts to include sample sizes larger than the required minimum will ensure that attrition among students does not affect the study's power.

Example #8: Complex sample with student assignment (fixed effects, with covariates)

Design: After recruiting all 20 schools and 50 study participants from each school, 25 students in each school are assigned to the program and 25 students in each school are assigned as controls. In PowerUp, this is noted as a Two-level Fixed Effects Blocked Individual Random Assignment Design, and we wish to compute the minimum detectable effect size (MDES) for this sample and compare this to the expected effect size of .20. We use 1 covariate at the student level with an R-square of $.7^2$=.49.

Note: Using a fixed effect will increase power relative to random effects, but this limits the generalizability of the findings to just the study participants. Random effects, which have lower power, allow generalizations beyond study participants.

Model 2.2: MDES Calculator for 2-Level Fixed Effects Blocked Individual Random Assignment Designs (BIRA2_1f)— School Intercepts and Interactions with TREATMENT		
Assumptions		**Comments**
Alpha Level (α)	0.05	Probability of a Type I error
Two-tailed or One-tailed Test?	2	
Power (1-β)	0.80	Statistical power (1-probability of a Type II error)
P	0.50	Proportion of Level 1 units randomized to treatment: $n_T / (n_T + n_C)$
R_1^2	0.49	Proportion of variance in Level 1 outcome explained by Block and Level 1 covariates
g*	1	Number of Level 1 covariates
n (Average Block Size)	50	Mean number of Level 1 units per Level 2 cluster (harmonic mean recommended)
J (Sample Size [# of Blocks])	20	Number of Level 2 units
M (Multiplier)	2.80	Computed from T_1 and T_2
T_1 (Precision)	1.96	Determined from alpha level, given two-tailed or one-tailed test
T_2 (Power)	0.84	Determined from given power level
MDES	**0.127**	Minimum Detectable Effect Size

Sample Language for Describing Results of Power Analyses:

A power analysis was calculated to ensure that expected sample sizes adequately power the study to detect significant differences between the program students and a comparison group of students who will not participate in the program. The program typically is able to serve about 500 students each program year, so the study plans to form two evenly sized groups of 500 program students and 500 comparison students using random assignment. Within each of the 20 schools participating in the study, 25 students will be assigned to the program and 25 students to the control group. A Two-level Fixed Effects Blocked Individual Random Assignment Design in PowerUp was used to compute the minimum detectable effect size (MDES) associated with this sample, assuming 1 covariate with an R-square of $.7^2$=.49. Based on the literature on math interventions provided to similar populations, an effect size of .20 is expected.

The power analysis (assuming an alpha of .05 and power of .80) reports that a sample size of 1,000 total students, or approximately 500 students per group, is adequate to detect an effect size of at least 0.13

compared to the expected effect size of .20. Therefore, the program sample estimates confirm that the available pool of students to include in the program and comparison groups appears to be sufficient for meeting the minimum sample size requirements for properly powering the study. The study also estimates a low attrition rate of 10% based on student attrition levels reported by the program over the past few years. Efforts to include sample sizes larger than the required minimum will ensure that attrition among students does not affect the study's power.