Paper #3 Abstract

Title

PowerUp!-Moderator: A Software Assisting the Design of Cluster Randomized Trials to Detect the Moderator Effects

Authors

Nianbo Dong, University of Missouri
Jessaca Spybrook, Western Michigan University
Ben Kelcey, University of Cincinnati

Background

Policy makers and researchers are not only interested in the program's main effects ("what works"), but also moderation effects ("work for whom, under what conditions"). A critical step in designing Cluster Randomized Trials (CRTs) to detect these effects is conducting a priori statistical power analyses. Statistical power for main effects in CRTs has an extensive research base and user-friendly computer programs for executing power analyses (e.g., Optimal Design, [Raudenbush, Spybrook, Congdon, Liu, & Martinez, 2011], CRT-Power [Borenstein & Hedges, 2012], and PowerUp! [Dong & Maynard, 2013]). However, research on power analysis to detect moderator effects in CRTs is very limited although more recently, some advancements have been made in this area (Bloom, 2005; Bloom & Spybrook, 2015; Jaciw, 2014; Mathieu, Aguinis, Culpepper, & Chen, 2012; Spybrook, Kelcey, & Dong, 2016).

Purpose

The purpose of this paper is to present results of recent advances in power analyses to detect the moderator effects in CRTs. This paper focus on demonstration of the software PowerUp!-Moderator (Dong, Kelcey, Spybrook, & Maynard, 2016), which was based on recent work (Dong, Kelcey, & Spybrook, 2016; Dong, Spybrook, & Kelcey, 2016; Spybrook, Kelcey, & Dong, 2016). This paper provides a resource for researchers seeking to design CRTs with adequate power to detect the moderator effects of the programs.

Methods

We implemented the closed-form power formulas that we previously derived and validated through simulations in an easy to use interactive Excel spreadsheet PowerUp!-Moderator. We first outline the software PowerUp!-Moderator. We then delineate the factors that are necessary for power analysis of moderator effects. Finally we demonstrate how to calculate the minimum detectable effect size difference (MDESD) and power using several examples.

Results

Table 1 covers the models and corresponding worksheets for calculation of minimum detectable effect size difference (MDESD) and power of moderator effects in two- and three-level cluster randomized trials (CRTs) (i.e., cluster random assignment designs[CRA]). Column 1 indicates the number of total levels of clustering (2 for two-level CRTs/CRA). Column 2 indicates the model number, e.g., CRA2-1N is for the model with a level-1 moderator (Colum 3) with nonrandomly varying slope (Column 4), CRA2-1R is for the model with a level-1 moderator (Colum 3) with random slope (Column 4), CRA2-2 is for the model with a level-2 moderator (Colum 3). Columns 5 and 6 contain the worksheet labels for the calculation of MDESD and power for the binary moderators, while Columns 7 and 8 contain the worksheet labels for the calculation of MDESD and power for the continuous moderators. The users can click these worksheet labels to go to the corresponding worksheets to conduct power analyses.

As in the power analysis of the main treatment effect, the power of the moderator effect in two-level CRTs is associated with the Type I error rate (α), one-tailed or two-tailed test, the unconditional intraclass correlation (ICC), the proportion of clusters in the treatment group (P), the proportion of variance explained by covariates, and the sample sizes for clusters and individuals. In addition, if the moderator is a binary variable, the power is also associated with the proportion of sample in one moderator subgroup; if the moderator is at level-1 with a random slope, the power is also associated with the effect heterogeneity (ω) for the level-1 moderator across level-2 units. We demonstrate the calculation of MDESD and power using several examples below.

Suppose a team of researchers are designing a two-level CRT to test the efficacy of a school-based intervention on student achievement. They are interested in student-level moderator effects and school-level moderator effects. The moderator can be a binary or continuous variable. They approach the moderator power analyses from two perspectives: (1) what is the MDESD given power of 0.80 and (2) what is the power for an effect size of 0.20. Table 2 shows the results of MDESD and power for the total numbers (J) of schools of 40 and 80 under some assumptions (In the presentation we will share resources with participants concerning the identification and selection of practical values of design parameters within a few different substantive examples.).

We have the following findings from Table 2. First, a design always has a smaller MDESD, or larger power for a fixed effect size when the level-2 sample size is bigger (e.g., J = 80 vs. 40). Second, the MDESD is larger, or the power is smaller for a fixed effect size when the moderator is at school level compared to the student level. Third, under these assumptions, when the moderator is at student level, the nonrandomly varying moderator slope model has a smaller MDESD, or bigger power for a fixed effect size than the random moderator slope model. Finally, the MDESD as defined by Cohen's *d* for the binary moderator is always twice the value of the MDESD defined by the standardized coefficient for the continuous moderator when the moderator is school level or the moderator is at student level with the nonrandomly varying slope.

Tables 3, 4, and 5 below demonstrate the calculation of MDESD and power for three examples under the same assumptions in Table 2 using the software that we developed. Table 4 demonstrates the calculation of MDESD regarding Cohen's *d* and 95% confidence interval for a binary level-1 moderator with nonrandomly varying slope. Table 5 demonstrates the calculation of MDESD regarding Cohen's *d* and 95% confidence interval for a continuous level-1 moderator with random slope. Table 6 demonstrates the calculation of power for a binary level-2 moderator. In all worksheets, the user needs to input the assumptions about the design parameters (e.g., ICC and R^2) and significance test (e.g., alpha level and one/two-tailed test) that are highlighted in yellow, and the results of the MDESD and its confidence interval, or power will be calculated automatically.

References

Bloom, H. S. (2005). Randomizing groups to evaluate place-based programs. In Howard S. Bloom (editor), *Learning more from social experiments: Evolving analytic approaches*, 115-172, New York: Russell Sage Foundation.

Bloom, H., & Spybrook, J. (2015). *Analysis of Statistical Precision for Multi-site Randomized Trials*. Paper presented at the 2015 Meeting for the Association for Public Policy and Management.

Borenstein, M. & Hedges, L. V. (2012). CRT-Power – Power Analysis for Cluster-Randomized and Multi-Site Studies [Computer software]. Englewood, NJ: Biostat.

Dong, N., & Maynard, R. A. (2013). 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, 6(1), 24-67*. doi: 10.1080/19345747.2012.673143.

Dong, N., Kelcey, B., Spybrook, J., & Maynard, R. A. (2016). PowerUp!-Moderator: A tool for calculating statistical power and minimum detectable effect size of the moderator effects in cluster randomized trials (Version 1.06) [Software]. Available from http://www.causalevaluation.org/

Dong, N., Spybrook, J., & Kelcey, B. (2016). *Power analyses for moderator effects with (non)random slopes in cluster randomized trials*. Manuscript submitted for publication.

Dong, N., Kelcey, B., & Spybrook, J. (2016). *Power analyses for moderator effects in three-level cluster randomized trials*. Manuscript submitted for publication.

Jaciw, A. (2014). An empirical study of design parameters for assessing the differential impacts for students in group randomized trials. Working paper http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2516005

Mathieu, J .E., Aguinis, H., Culpepper, S. A., and Chen, G. (2012). Understanding and estimating the power to detect cross-level interaction effects in multilevel modeling. *Journal of Applied Psychology, 97* (5), 951-966.

Raudenbush, S. W., Spybrook, J., Congdon, R., Liu, X, & Martinez, A. (2011). Optimal Design Software for Multi-level and Longitudinal Research (Version 2.01) [Computer software]. www.wtgrantfoundation.org.

Spybrook, J., Kelcey, B., & Dong, N. (2016). Power for detecting treatment by moderator effects in two and three-level cluster randomized trials. *Journal of Educational and Behavioral Statistics*. doi: 10.3102/107699861665544

Tables & Figures

Table 1. PowerUp!-Moderator to Detect Moderator Effects in 2- and 3-Level CRTs: Models and Corresponding

PowerUp!-Moderator to Detect Moderator Effects in 2- and 3-Level CRTs: Models and Corresponding Worksheets

	1	2	3	4	5	6	7	8
	Number of Total Levels of Clustering	Model Number	Level of Moderator	Slope of Lower Level Moderator	Binary Moderator		Continuous Moderator	
Study Design					MDESD Calculation	Power Calculation	MDESD Calculation	Power Calculation
Cluster Random Assignment Designs (Level of Assignment $\neq$ Level of Analysis)								
Simple Cluster Random Assignment (CRA), or Cluster Randomized Trials (CRTs)	2	CRA2-1N	1	Nonrandomly Varying	CRA2_1N_MDESD	CRA2_1N_Power	CRA2_1NC_MDESD	CRA2_1NC_Power
		CRA2-1R	1	Random	CRA2_1R_MDESD	CRA2_1R_Power	CRA2_1RC_MDESD	CRA2_1RC_Power
		CRA2-2	2	NA	CRA2_2_MDESD	CRA2_2_Power	CRA2_2C_MDESD	CRA2_2C_Power
	3	CRA3-1N	1	Nonrandomly Varying	CRA3_1N_MDESD	CRA3_1N_Power	CRA3_1NC_MDESD	CRA3_1NC_Power
		CRA3-1R	1	Random	CRA3_1R_MDESD	CRA3_1R_Power	CRA3_1RC_MDESD	CRA3_1RC_Power
		CRA3-2N	2	Nonrandomly Varying	CRA3_2N_MDESD	CRA3_2N_Power	CRA3_2NC_MDESD	CRA3_2NC_Power
		CRA3-2R	2	Random	CRA3_2R_MDESD	CRA3_2R_Power	CRA3_2RC_MDESD	CRA3_2RC_Power
		CRA3-3	3	NA	CRA3_3_MDESD	CRA3_3_Power	CRA3_3C_MDESD	CRA3_3C_Power

Table 2. MDESD and statistical power of two-level CRTs

Level of Moderator	Slope of Lower Level Moderator	MDESD				Power			
		Binary Moderator		Continuous Moderator		Binary Moderator		Continuous Moderator	
		$J = 40$	$J = 80$	$J = 40$	$J = 80$	$J = 40$	$J = 80$	$J = 40$	$J = 80$
1	Nonrandomly Varying	0.11	0.08	0.06	0.04	1.00	1.00	1.00	1.00
1	Random	0.26	0.18	0.25	0.17	0.56	0.86	0.63	0.91
2	NA	0.67	0.45	0.34	0.23	0.13	0.24	0.39	0.70

Note. Under the assumptions: $n = 100$, $\rho = 0.23$, $R_1^2 = 0.5$, $R_2^2 = 0.5$, $R_{2T}^2 = 0$, $\omega = 0.3$, $P = 0.5$, $Q = 0.5$, power = 0.8 for the calculation of MDESD, and effect size difference = 0.2 for the calculation of power, a two-sided test with $\alpha = 0.05$.

Table 3: MDESD and 95% confidence interval for a binary level-1 moderator with nonrandomly varying slope in a two-level CRT

Model CRA2-1N: MDESD Calculator for Two-Level Cluster Random Assignment Design — Treatment at Level 2 and Binary Moderator at Level 1 (Nonrandomly varying moderator slope model)

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)	0.23	Proportion of variance in outcome that is between clusters
P	0.50	Proportion of Level 2 units randomized to treatment: $J_T / (J_T + J_C)$
Q	0.50	Proportion of Level 1 units in Moderator subgroup: $n_1 / (n_1 + n_0)$
R_1^2	0.50	Proportion of variance in Level 1 outcomes explained by Level 1 covariates
g*	1	Number of Level 1 covariates excluding the moderator
n (Average Cluster Size)	100	Mean number of Level 1 units per Level 2 cluster (harmonic mean recommended)
J (Sample Size [# of Clusters])	40	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
MDESD	**0.110**	Minimum Detectable Effect Size Difference regarding Cohen's d.
95% Confidence Interval	**(0.033, 0.187)**	95% Confidence Interval of MDESD

Table 4: MDESD and 95% confidence interval for a continuous level-1 moderator with random slope in a two-level CRT

Model CRA2-1RC: MDESD Calculator for Two-Level Cluster Random Assignment Design — Treatment at Level 2 and Continuous Moderator at Level 1 (Random moderator slope model)		
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)	0.23	Proportion of variance in outcome that is between clusters
ω	0.30	The effect heterogeneity for the level-1 moderator across clusters in the model that is not conditional on treatment variable, which is the proportion of the variance between clusters on the effect of the moderator to the between-cluster residual variance. $\omega=\tau_{11}^2/\tau_{00}^2$
P	0.50	Proportion of Level 2 units randomized to treatment: J_T / ($J_T + J_C$)
R_1^2	0.50	Proportion of variance in Level 1 outcomes explained by Level 1 covariates
R_{2T}^2	0.00	Proportion of variance between Level-2 clusters on the effect of Level-1 moderator explained by level-2 predictors.
n (Average Cluster Size)	100	Mean number of Level 1 units per Level 2 cluster (harmonic mean recommended)
J (Sample Size [# of Clusters])	40	Number of Level 2 units
M (Multiplier)	2.88	Computed from T_1 and T_2
T_1 (Precision)	2.02	Determined from alpha level, given two-tailed or one-tailed test
T_2 (Power)	0.85	Determined from given power level
MDESD	**0.245**	Minimum Detectable Effect Size Difference regarding standardized coefficient
95% Confidence Interval	**(0.073, 0.418)**	95% Confidence Interval of MDESD

Table 5: Power for a binary level-2 moderator in a two-level CRT

Model CRA2-2: Power Calculator for Two-Level Cluster Random Assignment Design — Treatment at Level 2 and Binary Moderator at Level 2		
Assumptions		**Comments**
Alpha Level (α)	0.05	Probability of a Type I error
Two-tailed or One-tailed Test?	2	
Effect Size Difference	0.2	Effect Size Difference regarding Cohen's d.
Rho (ICC)	0.23	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)
Q	0.50	Proportion of Level 2 units in Moderator subgroup: J_1 / (J_1 + J_0)
R_1^2	0.50	Proportion of variance in Level 1 outcomes explained by Level 1 covariates
R_2^2	0.50	Proportion of variance in Level 2 outcome explained by Level 2 covariates
g*	1	Number of Level 2 covariates excluding the moderator and moderator*Treatment
n (Average Cluster Size)	100	Mean number of Level 1 units per Level 2 cluster (harmonic mean recommended)
J (Sample Size [# of Clusters])	80	Number of Level 2 units
Noncentrality Parameter	1.26	Automatically computed from the above assumptions
Power (1-β)	**0.236**	Statistical power (1-probability of a Type II error)