Power Calculations for Randomized Controlled Trials with Auxiliary Observational Data

Jaylin Lowe[1], Charlotte Z. Mann[1], Jiaying Wang[2], Adam Sales[3], and Johann Gagnon-Bartsch[1]

[1]University of Michigan [2]University of Southern California [3]Worcester Polytechnic Institute

Background

Randomized controlled trials (RCTs) have long been considered the "gold standard" in causal inference; however, they are often small and lead to estimates with high variance. Recent methods have sought to improve precision in RCTs by utilizing data from large observational datasets for covariate adjustment [1].

If successful, such methods can improve precision and therefore decrease the sample size necessary to achieve the desired power. However, the magnitude of the decrease is entirely dependent on how predictive a model trained on the auxiliary data will be for observations on some future, currently unknown, RCT data.

Our contribution: We present a method to determine a range of reasonable sample sizes to achieve the desired power when auxiliary data will be used for covariate adjustment.

Using Auxiliary Information For Covariate Adjustment

(1) Train a model on the auxiliary model.

(2) Use this model to make an auxiliary prediction for every observation in the RCT, such that potential outcomes [3, 4] for observation i are independent of its treatment assignment.

(3) Generate predicted potential observations for each observation in the RCT using the auxiliary prediction and refitting on the control units and the treatment units separately.

Decrease in sample size using this method is determined by how predictive the auxiliary model is for the RCT data. Our power calculation method aims at estimating how useful the auxiliary model will be for the RCT data, before the RCT data is obtained.

Our Method

(1) Break the auxiliary data up into subgroups.

(2) For each subgroup:

(2.1) Treat each subgroup as an RCT and the rest of the data as the auxiliary data.

(2.2) Calculate required sample size under this framework for each subgroup by obtaining out-of-bag predictions from a random forest run on the entire dataset. Use the residuals of these predictions as the estimate of σ^2

(3) Choose a sample size based on the overall range of sample sizes obtained from different subgroups.

Power Calculations

Typically, power calculations are done using the following formula [7]:

$$n = 2\sigma^2 \frac{(\xi_{1-\alpha/2} + \xi_{1-\beta})^2}{\Delta_A^2}$$

$\xi_{1-\alpha/2}$, $\xi_{1-\beta}$, and Δ_A are parameters, usually set to 0.05, 0.20, and 20% of the standard deviation of the outcome in the population.

σ^2 is the true variance of the outcome in the population, typically replaced by an estimate of the variance of the outcome obtained from a sample. We replace it with the variance of the residuals from the out of bag predictions from a random forest.

Subgroup Formation

Method 1: Covariate Value

Subgroups can be formed by dividing by the values of covariates. For numerical covariates, we recommend dividing into 10 approximately equally sized groups. For categorical covariates, we recommend creating a subgroup for each level of the covariate.

Method 2: Best/Worst Case Scenario

We group observations according to the following process:

(1) Fit an initial random forest on the auxiliary dataset. Obtain the out of bag predictions.

(2) Calculate the absolute value of the error of these predictions.

(3) Fit a second random forest where the outcome is the absolute value of the errors. Call out-of-bag predictions from this model the "predicted error".

(4) Split observations into groups based on the predicted error.

This groups observations according to how predictive we would expect the auxiliary model to be for that group. If an observation has high predicted error, that means that the other observations in the auxiliary data are not helpful in predicting outcomes for an observation with those covariate values—the worst case scenario. Conversely, if the auxiliary model performs similarly on the RCT data as it does on observations with low predicted error, then utilizing the auxiliary data should improve precision in the RCT estimates—the best case scenario.

Example: Cognitive Tutor Algebra Study

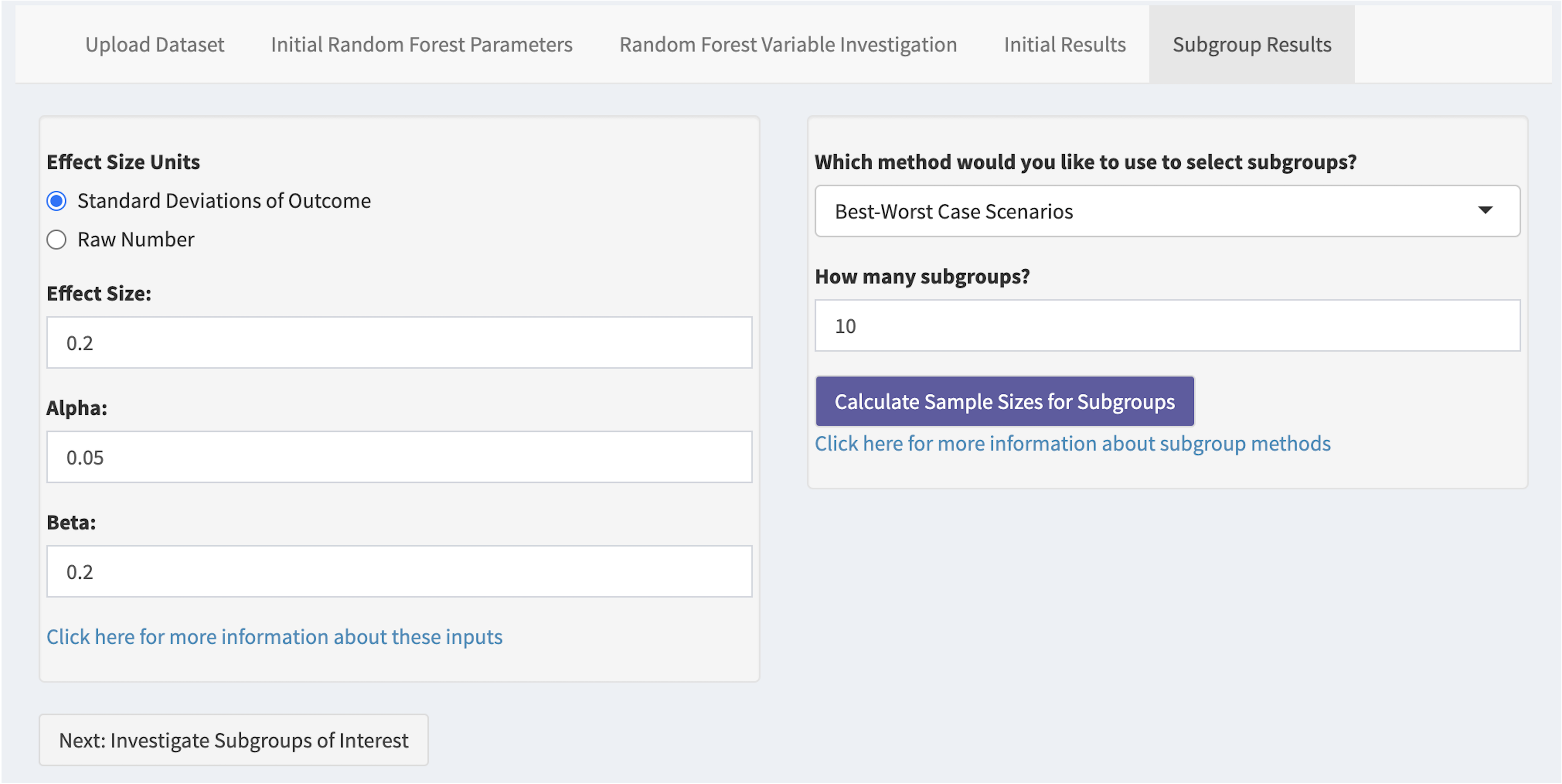
As an example, we show how these methods can be applied in practice, using an efficacy trial for the Cognitive Tutor Algebra I (CTAI) curriculum. CTAI was a new technology based algebra curriculum with personalized automated tutoring software [2]. Schools were randomised to either implement CTAI (treatment) or use their standard curriculum (control) and the groups were compared using subsequent mathematics test scores.

Standardized test scores and other demographic information is publicly available for all schools in Texas[5, 6]. We started by creating subgroups using the best/worst case scenario method. The results are shown in the table below.

	Auxiliary Data?	
Decile	Yes	No
1	19	92
2	47	200
3	50	272
4	81	424
5	108	382
6	113	516
7	123	497
8	191	622
9	370	1094
10	998	1829

Graphical User Interface

A Shiny app aimed at helping users implement these methods can be found at github.com/jaylinlowe/dRCT-power. Researchers can create their own subgroups using any of the three methods, adjust the other parameters in the formula, and compare the resulting sample sizes. In addition, the app contains features aimed at helping researchers determine which subgroups to investigate.



Acknowledgments

The research reported here was supported by the Institute of Education Sciences, U.S. Department of Education, through Grant R305D210031. The opinions expressed are those of the authors and do not represent views of the Institute or the U.S. Department of Education.

References

[1] J. A. Gagnon-Bartsch, A. C. Sales, E. Wu, A. F. Botelho, J. A. Erickson, L. W. Miratrix, and N. T. Heffernan. Precise unbiased estimation in randomized experiments using auxiliary observational data. *Journal of Causal Inference*, 11(1):20220011, Jan. 2023. Publisher: De Gruyter.

[2] J. F. Pane, B. A. Griffin, D. F. McCaffrey, and R. Karam. Effectiveness of Cognitive Tutor Algebra I at Scale. *Educational Evaluation and Policy Analysis*, 36(2):127–144, 2014. Publisher: [American Educational Research Association, Sage Publications, Inc.].

[3] D. B. Rubin. Estimating causal effects of treatments in randomized and nonrandomized studies. *Journal of Educational Psychology*, 66(5):688–701, 1974. Place: US Publisher: American Psychological Association.

[4] J. Splawa-Neyman, D. M. Dabrowska, and T. P. Speed. On the Application of Probability Theory to Agricultural Experiments. Essay on Principles. Section 9. *Statistical Science*, 5(4):465–472, 1923.

[5] Texas Education Agency - Academic Excellence Indicator System. Access at: https://rptsvr1.tea.texas.gov/perfreport/aeis/index.html.

[6] Texas Assessment of Knowledge and Skills (TAKS), 2017. Access at: https://tea.texas.gov/student-assessment/testing/student-assessment-overview/2017-ig-taks.pdf.

[7] J. Wittes. Sample Size Calculations for Randomized Controlled Trials. *Epidemiologic Reviews*, 24(1):39–53, July 2002.

Poster created using template by David Gamba, University of Michigan. Available at `https://www.overleaf.com/latex/templates/university-of-michigan-umich-poster-template/xpnqzzxwbjzc`.