Online Appendix A: A Simplified Power Analysis

for "Data Issues in Developing Valid ROI/ROR Estimates"
Journal of Rehabilitation Administration (2019, forthcoming)

Steven Stern
Stony Brook University

Chris Clapp
University of Chicago

John Pepper
University of Virginia

Robert Schmidt
University of Richmond

November 2018

Consider a simple example where the outcome of interest is binary employment and the service is also binary (either one receives the service or not). Let x_{1i} be a random employment outcome for person i=1,2,..,n_1 from the subsample receiving no service and x_{2i} be one from the subsample receiving service, i=1,2,..,n_2. Define p_{ji} as the probability that person i from subsample j is employed. Let $\overline{x}_j$ be the average outcome for subsample j=1,2 and $\overline{p}_j$ be the average probability for people from subsample j. Then the t-statistic for a test of H_0:p_1=p_2 is[1]

$$(1) \qquad t_n = \frac{\overline{x}_1 - \overline{x}_2}{\sqrt{\sum_j \left[\frac{1}{n_j}\overline{p}_j\left(1-\overline{p}_j\right) + \frac{1}{n_j^2}\sum_i\left(p_{ij}-\overline{p}_j\right)^2\right]}}$$

Consider an example similar to those in Dean et al. (2015, 2017, 2018, 2019) where $\overline{p}_1$-$\overline{p}_2$=$\overline{x}_1$-$\overline{x}_{j2}$=Δ, 20% of the sample receives the service of interest, and the variance of probabilities within each subsample is

$$\frac{1}{n_j}\sum_i\left(p_{ij} - \overline{p}_j\right)^2 = 0.09.$$

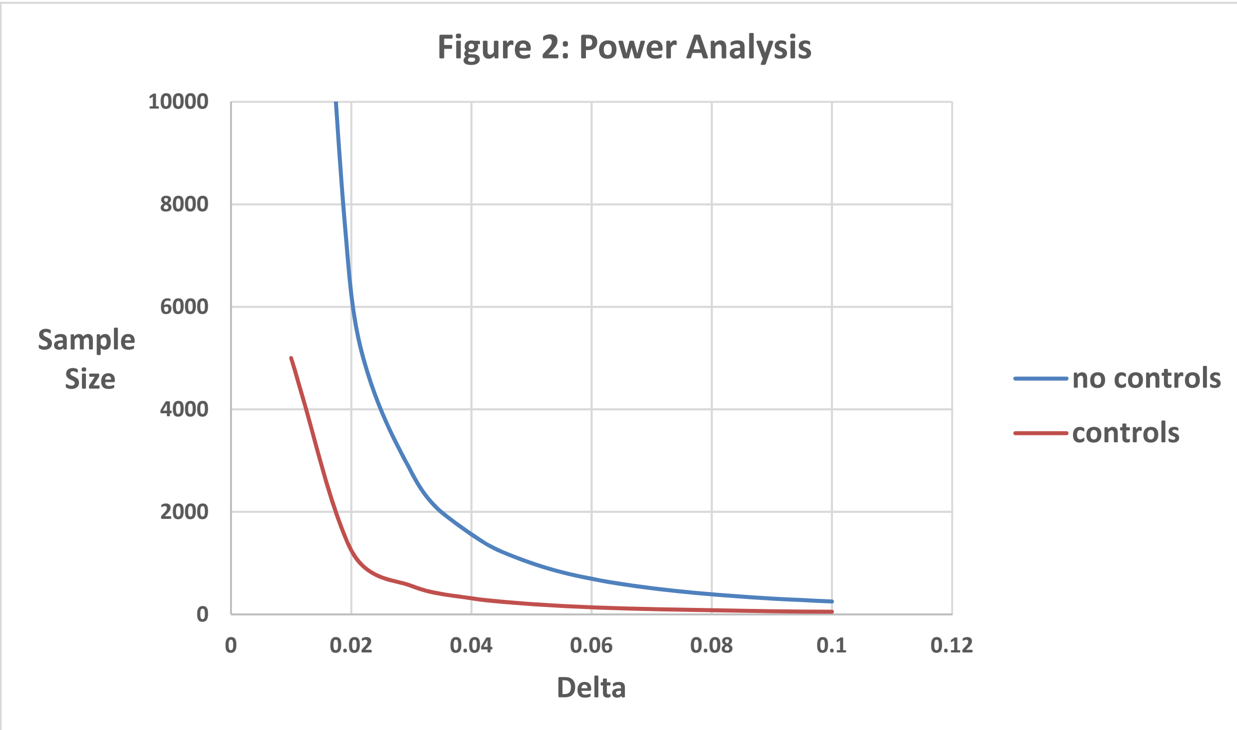
Then, using $_{tn}$}=2 as a critical value for statistical significance, we get combinations of n=n_1+n_2 and Δ as n=2.5/Δ^2 as represented by the dashed line in Figure 2. For example, if the true value of is .05, then only 1000 observations are needed to get a statistically significant estimate; but, if the true value of is 0.02, then 6250 observations are needed. A large source of the randomness is caused by not controlling for variation in employment probabilities within each subsample likely

---

[1] This follows from

$$\sqrt{n}(\overline{x}_j - \overline{p}_j) \sim N(0,\sigma_j^2)$$
$$\sigma_j^2 = \overline{p}_j \ (1-\overline{p}_j\ ) + (1/(n_j\}))\sum(p_{ij}\} - \overline{p}_j)^2.$$

caused by other observable characteristics of each person (see Section 4 of the paper). If we control for that variation, then the last term in the denominator in equation (1) disappears, and the new relationship between n and Δ is $n=0.5/\Delta^2$ as represented by the solid line in Figure 2. Now a sample of size 2000 would provide statistically significant estimates even if $\Delta=0.016$.



The power analysis problems in most estimation problems are more complex than the straightforward one done here. Complications are due to multiple services (Dean et al., 2015, 2017, 2018a, 2018b), multiple outcome measures (e.g., employment and earnings), and the need for more complex econometric methods required by endogenous service receipt (Heckman, Lalonde, and Smith, 1999; Aakvik, Heckman, and Vytlacil, 2005; Dean et al., 2015, 2017, 2018a, 2108b).

References

Aakvik, A., Heckman, J., & Vytlacil, E. (2005). Estimating treatment effects for discrete outcomes when responses to treatment vary: An application to Norwegian vocational rehabilitation programs. Journal of Econometrics, 125, 15-51.

Dean, D. H., Pepper, J. V., Schmidt, R. M., & Stern S. N. (2015). The effects of vocational rehabilitation for people with cognitive impairments. International Economic Review, 56(2), 399-426.

Dean, D. H., Pepper, J. V., Schmidt, R. M., & Stern, S. N. (2017). The effects of vocational rehabilitation for people with mental illness. Journal of Human Resources, 52(3), 826-858.

Dean, D. H., Pepper, J. V., Schmidt, R. M., & Stern, S. N. (2018). The effects of vocational rehabilitation for people with physical impairments. Journal of Human Capital, 12(1), 1-37.

Dean, D. H., Pepper, J. V., Schmidt, R. M., & Stern, S. N. (2019). The effects of youth transition programs on labor market outcomes. Economics of Education Review, 68 (February), 68-88.

Heckman, J., LaLonde, R., and Smith, J. (1999). The economics and econometrics of active labor market programs. Handbook of Labor Economics Volume 3. (eds.) O. Ashenfelter and D. Card. Amsterdam: North-Holland,1865-2097.