



**REFLECTIONS  
ON  
METHODOLOGY**

**JULY 2018**

**NEW YORK**

16 East 34th Street  
New York, NY 10016  
Tel: 212 532 3200

**OAKLAND**

475 14th Street  
Suite 750  
Oakland, CA 94612  
Tel: 510 663 6372

**WASHINGTON, DC**

1990 M Street, NW  
Suite 340  
Washington, DC 20036

**LOS ANGELES**

11965 Venice Boulevard  
Suite 402  
Los Angeles, CA 90066

www.mdrc.org



# Addressing Methodological Challenges in Naturally Occurring Lotteries

By *Rebecca Unterman*

*This post is one in a series highlighting MDRC's methodological work. Contributors discuss the refinement and practical use of research methods being employed across our organization.*

In a [previous post](#), I described how school enrollment processes that contain naturally occurring lotteries provide researchers with exciting opportunities to learn about the effects of policies and programs. Because such natural lotteries randomly determine which schools students are assigned to, there should be no measurable or unmeasurable differences between the students who win and those who lose the opportunity to attend the school; therefore, any differences in the students' educational outcomes can be attributed to that opportunity. This allows researchers to identify the causal effect of the school model on students. In this follow-up post, I present a few methodological issues common to lottery-based analyses — constrained statistical power, imperfect compliance, and restricted generalizability — and briefly discuss how they can be addressed.

## CONSTRAINED STATISTICAL POWER

As in a standard random assignment design, a lottery-based sample's statistical power is a function of the number of random assignment blocks (lotteries), the number of students per block, the number of covariates, the predictive power of the blocks and covariates for the relevant student outcome ( $R^2$ ), and the proportion of the sample assigned to treatment. Yet in reality, it may be a bit more complicated, for at least two reasons. First, there are often many lottery blocks containing just a few students, which use up degrees of freedom in the calculation. If those blocks are not sufficiently predictive of the relevant student outcome to counterbalance that loss, the resulting sample could have less statistical power than anticipated. Second, lotteries can result in most students being in the winning category (if a school is barely oversubscribed) or most students being in the losing category (if a large number of students compete for a small number of seats). Both situations result in lower statistical power than if half the applicants were winners in each school. Researchers can [learn more](#) and easily see the effects of these two components by entering their data in the PowerUp! tool available [online](#).<sup>1</sup>

## IMPERFECT COMPLIANCE

Compliance with lottery assignment is likely to be imperfect: A student who wins admission might not attend the school, and a student who does not win admission might attend anyway via subsequent steps in the school assignment process (or by showing up at the school on the first day of classes). While noncompliance may also occur in traditional random assignment studies, with naturally occurring school lotteries it occurs more often due to the often complex, multistep nature of the school assignment process.

In the case of noncompliance it is often useful to estimate the effect of enrolling in the school in addition to estimating the effect of winning the lottery. This can be done using a standard appli-

---

<sup>1</sup>Because the size of the priority blocks may vary greatly, it is advisable to use the harmonic mean group size when estimating the minimum detectable effect size (MDES) rather than the arithmetic mean.

cation of instrumental variables analysis,<sup>2</sup> an approach often applied in randomized experiments and lottery-based studies.<sup>3</sup> This analysis requires the researcher to make a key assumption called the “exclusion restriction,” in which assignment to the school affects a student’s future outcomes *only through enrollment in the school*. For example, if a student won a seat in the school and as a result learned about and signed up for additional after-school opportunities and summer activities, assignment to the school could affect the student’s future outcomes through multiple pathways. Therefore, a simple model in which school assignment affects the student only through school enrollment may not be accurate. (However, a team of researchers, including a few senior MDRC methodologists, are currently working on a methodological paper that will explain how the exclusion restriction may be relaxed under certain circumstances when conducting multisite instrumental variables analyses.)

### RESTRICTED GENERALIZABILITY

The generalizability of estimates from the lottery sample may be constrained for a few reasons related to lottery identification. First, if only a few study schools are oversubscribed, this may be a sign that these schools are different from other study schools. The oversubscribed schools may be more popular because they are of higher quality than the others, or they may have a better relationship to the community. The results from an analysis of the opportunity to attend these schools will be important, but they cannot be generalized to all schools in operation during the study period. Second, if it is just an admissions priority group within a school that is oversubscribed, it may be that students in that priority group are different from the other students attending the school. For example, if the school offers preference first to siblings of current students and second to all other students, and the lottery occurs within the group of all other students, the lottery-based analysis may not accurately capture the school’s effect on siblings. In this case the researcher should take care to note the proportion of each school’s student body represented by the oversubscribed priority group and be clear about whether the findings from this sample can be generalized to students from other admissions priority groups. (Likewise, the results of the instrumental variables analysis only apply to students who attend the school *because they won the lottery*; they do not apply to those students who would always find a way to enroll in the school, through later steps in the assignment process.)

In conclusion, while school lottery-based analyses provide researchers with exciting opportunities to study programs, they also require that researchers have a strong understanding of the data available and of the specific admissions processes involved.

---

<sup>2</sup>Angrist, Imbens, and Rubin (1996); Gennetian, Morris, Bos, and Bloom (2005).

<sup>3</sup>Abdulkadiroğlu et al. (2011); Bloom and Unterman (2014); Ludwig and Kling (2007).