

# Evaluating Education Programs That Have Lotteried Admission and Selective Attrition

John Engberg, *RAND Corporation*

Dennis Epple, *Carnegie Mellon University*

Jason Imbrogno, *Carnegie Mellon University*

Holger Sieg, *University of Pennsylvania*

Ron Zimmer, *Vanderbilt University*

We study the effectiveness of magnet programs in an urban district that ration excess demand by admission lotteries. Differential attrition arises since students who lose the lottery are more likely to pursue options outside the school district than students who win the lottery. When students leave the district, important outcome variables are often not observed. The treatment effects are not point-identified. We exploit known quantiles of the outcome distribution to construct informative bounds on treatment effects. We find that magnet programs improve behavioral outcomes but have no significant effect on achievement.

## I. Introduction

The purpose of this article is to study the effectiveness of educational programs that use lottery-based admission procedures. We focus on over-

We would like to thank Josh Angrist, Eric Bettinger, Moshe Buchinsky, Judy Cullen, Stefan Holderlein, Guido Imbens, Blaise Melly, Robert Moffitt, Parag

[*Journal of Labor Economics*, 2014, vol. 32, no. 1]  
© 2014 by The University of Chicago. All rights reserved.  
0734-306X/2014/3201-0002\$10.00

subscribed magnet programs in an urban district. While debates surrounding the effectiveness of other school choice options such as charter schools and educational vouchers have attracted much attention from researchers and policy makers, magnet programs have gotten less attention despite the fact that they are much more prevalent than charter schools or educational voucher programs. Most urban school districts typically operate a variety of magnet programs that are popular and, therefore, oversubscribed.

Many school districts use lotteries to determine access to oversubscribed educational programs. Lottery winners are accepted into the program, with the ultimate choice of attendance left to the student and his family. Lottery losers do not have the option to participate in the program but have many different outside options. As a consequence, lottery losers often decide to pursue options outside of the traditional public school system and attend charter or private schools. Educational outcomes are often not observed for students who leave the school system, which creates a missing data problem. If attrition rates differ by lottery status, the randomization inherent in the lottery assignment is not necessarily sufficient to identify meaningful treatment effects. However, we can still identify and estimate informative bounds on treatment effects under fairly weak assumptions about the nature of the attrition problem.<sup>1</sup>

Lottery-based admission can be viewed as an experimental design with multiple sources of noncompliance that arise from parental or student decisions. We focus on two of the most important outside options: parents can send their children to a nonmagnet program within the district or they can leave the school district and send their children to a private school or a public school in a different district. We model this behavior as noncompliance with the intended treatment using latent household types. Our approach builds on the work of Angrist, Imbens, and Rubin (1996) and adds two additional latent types to the framework to deal with multiple sources of noncompliance. Differential attrition arises in our framework if there exists a household type that complies with the lottery and participates in the program if it wins the lottery but leaves the district if it loses the lottery. We denote these households as “at risk” since they are at risk of leaving the district.

Our findings show that approximately 25% of applicants to magnet programs that serve elementary school students are at risk. That suggests that

---

Pathak, Chris Taber, Ken Wolpin, Jeff Wooldridge, Tiemen Woutersen, and participants at numerous conferences and seminars. We would also like to thank the “midsized urban school district” for sharing its data. Financial support for this research is provided by the Institute of Education Sciences (IES R305A070117 and R305D090016). Contact the corresponding author, Holger Sieg, at holgers@econ.upenn.edu. Data are available as supplementary material online.

<sup>1</sup> Selective attrition may also arise when lottery winners who initially participate in the program drop out because they experience unfavorable outcomes.

magnet programs help to attract and retain students. These households come from neighborhoods that have higher incomes and a higher fraction of more educated households than neighborhoods preferred by households that stay in the district regardless of the outcome of the lottery. The at-risk households have many options outside the public school system, but apparently they view the existing magnet programs as desirable programs for their children. The market for elementary school education is more competitive than the market for middle and high school education; that is, the fraction of households at risk declines with the age of the students.

Since the fraction of at-risk households is significantly larger than zero, we face a missing data problem. We do not observe educational outcomes when these households leave the district after they do not win the lottery. As a consequence, we cannot point-identify the causal effect of magnet programs on achievement or other behavioral outcomes.<sup>2</sup> We therefore focus the remainder of the analysis on estimating informative bounds on these treatment effects. One prominent approach that is often used in the partial identification literature relies on “worst-case” scenarios to construct bounds for treatment effects (Manski 1990). Horowitz and Manski (2000) provide a framework that exploits the assumption that the support of the outcome variable is bounded to deal with nonrandom attrition. We follow this approach but use known quantiles of the outcome distribution to create worst-case scenarios. In particular, we use the distribution of state test scores and the districtwide distribution of offenses, suspension days, tardies, and absences.<sup>3</sup> To our knowledge, we are the first to use these ideas to bound treatment effects of educational programs.

We find that our bounds analysis is informative and demonstrates that magnet programs offered by the district improve behavioral outcomes such as offenses, attendance, and timeliness. Our findings for achievement effects are mixed. While the point estimates of the upper and lower bounds point to positive treatment effects, sample sizes are still too small to provide precise estimates. This is largely the case because standardized achievement tests were conducted only in grades 5, 8, and 11 during most of our sample period.

The rest of the article is organized as follows. Section II provides a brief review of the literature. Section III discusses identification and estimation of treatment effects when program participation is partially determined by

<sup>2</sup> If there are two different types of compliers, the instrumental variable estimator does not identify a local average treatment effect. In a related paper, Heckman, Urzua, and Vytlačil (2006) also consider multiple unordered treatments with an instrument shifting agents into one of the treatments.

<sup>3</sup> We also implement Lee’s (2009) approach, which uses sample trimming rules to construct informative bounds. Note that this approach is closely related to that of Zhang and Rubin (2003).

lotteries and selective attrition cannot be ignored. Section IV provides some institutional background for our application and discusses our main data sources. Section V reports the empirical findings of our article. Finally, we offer some conclusions and discuss the policy implications of our work in Section VI.

## II. Literature Review

Our article is related to a growing literature that evaluates educational programs using lottery-based estimators.<sup>4</sup> Lotteries were used by Rouse (1998) to study the impact of the Milwaukee voucher program. Angrist et al. (2002) also study the effects of vouchers when there is randomization in selection of recipients from the pool of applicants using data from Colombia. Hoxby and Rockoff (2005) use lotteries to study Chicago charter schools. Cullen, Jacob, and Levitt (2006) have analyzed open-enrollment programs in the Chicago Public Schools. Hastings, Kane, and Staiger (2010) estimate a model of school choice based on stated preferences for schools in Charlotte, North Carolina. Since school attendance was partially the outcome of a lottery, they use the lottery outcomes as instruments to estimate the impact of attending the first-choice school. Abdulkadiroglu et al. (2009) and Hoxby and Murarka (2009) study charter schools in Boston and New York, respectively, and find strong achievement effects. Dobbie and Fryer (2009) study a social experiment in Harlem and show that high-quality schools or high-quality schools coupled with community investments generate the highest achievement gains. All of these papers focus on applications in which selective attrition is not present and therefore do not explicitly deal with the key selective attrition problem discussed in this article.<sup>5</sup>

Currently, there are a number of approaches that have been proposed in the econometric literature that deal with selection and attrition problems. Heckman (1974, 1979) explicitly models an outcome and a selection equation.<sup>6</sup> In some applications, there exists an exogenous variable that affects the selection but not the outcome equation. These types of exclusion restrictions can be used in both parametric and semiparametric estimation techniques.<sup>7</sup> In our application, there are no obvious exclusion restrictions.

When exclusion restrictions are not available, one can often construct bounds for the treatment effects. Horowitz and Manski (2000) provide a

<sup>4</sup> Angrist (1990) introduced the use of lotteries to study the impact of military service on earnings.

<sup>5</sup> Angrist et al. (2002) encounter a related issue of selective test participation since students in private schools are more likely to take college entrance exams than public school students.

<sup>6</sup> An alternative approach follows Rubin (1976) and assumes that data are missing at random, after conditioning on a set of observed variables.

<sup>7</sup> Some well-known examples are Heckman (1990), Ahn and Powell (1993), and Das, Newey, and Vella (2003).

general framework for dealing with nonrandom attrition that exploits the assumption that the support of the outcome variable is bounded. As we discussed in detail in the introduction, we follow this approach but use known quantiles of the outcome distribution to create bounds on the treatment effect.<sup>8</sup>

An alternative approach is based on the principal stratification method that is popular in the statistics literature (Frangakis and Rubin 2002). This approach classifies individuals in latent groups according to the joint values of the potential outcome variables. For example, in a standard selection model, there are four latent groups that are implicitly defined by the potential outcomes of the employment decision in the treated and untreated state.<sup>9</sup> Our approach also relies on latent types and builds on Angrist et al. (1996) to account for the multiple sources of noncompliance.

### III. Identification and Estimation

#### A. The Research Design

We consider a research design that arises when randomization determines eligibility to participate in an educational program. A parent has to decide whether or not to enroll a student in a magnet program offered by a school district.<sup>10</sup> We consider only households that participate in a lottery that determines access to an oversubscribed magnet program. Let  $W$  denote a discrete random variable that is equal to one if the student wins the lottery and zero if he loses. Let  $w$  denote the fraction of households that win the lottery.

We assume that a student who wins the lottery has three options: participate in the magnet program; participate in a different, nonmagnet program offered by the same school district; or leave the district and pursue educational opportunities outside the district. A student who loses and is not an always-taker has only the last two options. Let  $M$  be one if a student attends the magnet program and zero otherwise. Finally, let  $A$  denote a random variable that is one if a student attends a school in the district and zero otherwise.

To model compliance with the intended treatment, we define five latent types to classify households into compliers and noncompliers.<sup>11</sup>

<sup>8</sup> Blundell et al. (2007) develop bounds for the quantiles of the treatment distribution rather than using an extreme quantile of the outcome distribution to bound the average treatment effect.

<sup>9</sup> In a recent application, Barnard et al. (2003) study the effect of school choice on test scores, and Zhang, Rubin, and Mealli (2009) evaluate the Job Corps training program.

<sup>10</sup> We use the terms “parent” or “household” to describe the decision maker and “student” to describe the person who participates in the program.

<sup>11</sup> Appendix B discusses the assumptions needed to derive these five types from the 16 possible types.

## DEFINITION 1.

1. Let  $s_m$  denote the fraction of “complying stayers.” These households will remain in the district when they lose the lottery. If they win the lottery, they comply with the intended treatment and attend the magnet school.
2. Let  $s_n$  denote the fraction of “noncomplying stayers.” These households will remain in the district when they lose the lottery. If they win the lottery, they will not comply with the intended treatment and instead will attend a nonmagnet program in the district.
3. Let  $l$  denote the fraction of “leavers.” These are households that will leave the district regardless of whether they are admitted to the magnet program.<sup>12</sup>
4. Let  $r$  denote the fraction that is “at risk.” These households will remain in the district and attend the magnet program if admitted to the magnet program, and they will leave the district otherwise.
5. Let  $a_t$  denote the fraction of “always-takers.” They will attend the magnet school regardless of the outcome of the lottery.

Since the household type is latent, one key empirical problem is identifying and estimating the proportions of each type in the underlying population. These parameters are informative about the effectiveness of magnet programs in attracting and retaining households that participate in the lottery. Moreover, we will show that households at risk cause the selective attrition problem.

The latent types of households are likely to differ in important characteristics, and we need to characterize these differences. If households at risk differ among observed characteristics from the other latent types, they may also differ by unobserved characteristics. As a consequence, ignoring the selective attrition problem will be problematic. By characterizing the observed characteristics of all latent types, we can thus gain some important insights into the potential importance of the selective attrition problem.

To formalize these ideas, consider a random vector  $X$  that measures observed household characteristics such as income or socioeconomic status.

<sup>12</sup> Parents have incomplete information and need to gather information to learn about the features of different programs. Parents have to sign up for lotteries months in advance. At that point, they have not accumulated all relevant information. Once they have accumulated all relevant information, they may decide to opt out of the public school system if their preferred choice dominates the program offered by the district. In addition, household circumstances may change. For example, parents may obtain a job that requires moving to a different metropolitan area. Note that there are typically no penalties for participating in the lottery and declining to participate in the program.

Appealing to our decomposition, let  $\mu_r$ ,  $\mu_{s_m}$ ,  $\mu_{s_n}$ ,  $\mu_l$ , and  $\mu_{a_t}$  denote the means of the random vector  $X$  conditional on belonging to group  $r$ ,  $s_m$ ,  $s_n$ ,  $l$ , and  $a_t$ , respectively. Below we discuss how to identify and estimate the parameters  $(w, r, s_n, s_m, l, a, \mu_r, \mu_{s_n}, \mu_{s_m}, \mu_l, \mu_{a_t})$ .

Let  $T$  be an outcome measure of interest, for example, the score on a standardized achievement test. Following Neyman (1923) and Fisher (1935), we adopt standard notation in the program evaluation literature and consider a model with three potential outcomes:

$$T = AMT_1 + A(1 - M)T_0 + (1 - A)T_2, \quad (1)$$

where  $T_1$  denotes the outcome if the student attends the magnet school,  $T_0$  if he attends a different program in the district, and  $T_2$  if he attends a school outside of the district.<sup>13</sup> We will later assume that  $T$  is not observed for students who do not attend a public school within the district; that is,  $T_2$  is not observed. This assumption is plausible since researchers typically have access to data from only one school district. Private schools rarely provide access to their confidential data and often do not administer the same standardized tests as public schools. Attention, therefore, focuses on the individual treatment effect  $\Delta = T_1 - T_0$ . Note that  $\Delta$  is unobserved for all students.

Conceptually, we can define five different average treatment effects, one for each latent group:<sup>14</sup>

$$ATE_{\text{Type}} = E[T_1 - T_0 | \text{Type} = 1], \quad \text{Type} \in \{s_n, s_m, R, L, A_t\}. \quad (2)$$

The key research question is then whether we can identify and estimate these types of treatment effects when selective attrition is important. To answer this question, we first discuss how to characterize the extent of the selective attrition problem. We then derive bounds estimators for the relevant treatment effects.

Finally, it is useful to compare our approach to the one developed in Angrist et al. (1996). Note that we have two types of “never-takers” that we denote by “noncomplying stayers” and “leavers.” Similarly, we have two types of “compliers” that we denote by “complying stayers” and “at-risk”

<sup>13</sup> This approach shares many similarities with the “switching regression” model introduced into economics by Quandt (1972), Heckman (1978, 1979), and Lee (1979). Heckman and Robb (1985) and Bjorklund and Moffitt (1987) treated heterogeneity in treatment as a random coefficients model. It is also known in the statistical literature as the Rubin model developed in Rubin (1974, 1978). See also Heckman and Vytalil (2007) for an overview of the program evaluation literature.

<sup>14</sup> There are other effects that may also be of interest such as treatment effect on the treated or the marginal treatment effect. For a discussion, see, among others, Heckman and Vytalil (2005) and Moffitt (2008).

households. The main difference arises because individuals have more than one outside option and outcomes are not observed for at-risk households that leave the district when they lose the lottery.

### B. Identification of the Fraction of Latent Types

First we need to establish the information set of the researcher. We observe probabilities and conditional means for the feasible outcomes shown in table 1. Note that only six of the eight outcomes listed in table 1 are possible since a student attending a magnet program ( $M = 1$ ) must also attend a public school ( $A = 1$ ).

Identification can be established sequentially. First, we discuss identification of the probabilities that characterize the shares of the latent types. We have the following result.

**PROPOSITION 1.** The parameters  $(w, r, s_n, s_m, l, a_t)$  are identified by the six nondegenerate probabilities in table 1.

*Proof.* Parameter  $w$  is the fraction that wins the lottery:

$$\begin{aligned} w = & \Pr(W = 1, M = 1, A = 1) + \Pr(W = 1, M = 1, A = 0) \\ & + \Pr(W = 1, M = 0, A = 1) + \Pr(W = 1, M = 0, A = 0). \end{aligned} \quad (3)$$

Given  $w$ ,  $s_n$  is identified from  $(1, 0, 1)$ :

$$s_n = \Pr(W = 1, M = 0, A = 1)/w; \quad (4)$$

$l$  is identified from  $(1, 0, 0)$ :

$$l = \Pr(W = 1, M = 0, A = 0)/w; \quad (5)$$

**Table 1**  
**Observed Outcomes**

	$W$	$M$	$A$	$\Pr\{W, M, A\}$	$\frac{E[X W, M, A]}{= E[X M, A]}$
1	1	1	1	$w(r + s_m + a_t)$	$\frac{r\mu_r + s_m\mu_{s_m} + a_t\mu_{a_t}}{r + s_m + a_t}$
2	1	1	0	Not possible	
3	1	0	1	$ws_n$	$\mu_{s_n}$
4	1	0	0	$wl$	$\mu_l$
5	0	1	1	$(1 - w)a_t$	$\mu_{a_t}$
6	0	1	0	Not possible	
7	0	0	1	$(1 - w)(s_n + s_m)$	$\frac{s_n\mu_{s_n} + s_m\mu_{s_m}}{s_n + s_m}$
8	0	0	0	$(1 - w)(r + l)$	$\frac{r\mu_r + l\mu_l}{r + l}$



$a_t$  is identified from  $(0, 1, 1)$ :

$$a_t = \Pr(W = 0, M = 1, A = 1)/(1 - w). \quad (6)$$

Given  $w$  and  $s_n$ ,  $s_m$  is identified from  $(0, 0, 1)$ :

$$s_m = \Pr(W = 0, M = 0, A = 1)/(1 - w) - s_n. \quad (7)$$

Given  $a_t$ ,  $l$ ,  $s_n$ , and  $s_m$ ,  $r$  is identified from the identity

$$r = 1 - l - s_m - s_n - a_t. \quad (8)$$

QED

Note that there is no overidentification at this stage since the six probabilities in table 1 add up to one, and the last three nondegenerate probabilities add up to  $1 - w$ .

Next we discuss identification of the five conditional means of household characteristics. We have the following result.

**PROPOSITION 2.** Given  $(w, r, s_n, s_m, l, a_t)$ , the parameters  $(\mu_r, \mu_{s_m}, \mu_{s_n}, \mu_l, \mu_{a_t})$  are identified by the observed conditional expectations in table 1.

*Proof.* Parameter  $\mu_l$  is identified from  $(1, 0, 0)$ :

$$\mu_l = E(X|W = 1, M = 0, A = 0). \quad (9)$$

Similarly,  $\mu_{s_n}$  is identified from  $(1, 0, 1)$ :

$$\mu_{s_n} = E(X|W = 1, M = 0, A = 1) \quad (10)$$

and  $\mu_{a_t}$  is identified from  $(0, 1, 1)$ :

$$\mu_{a_t} = E(X|W = 0, M = 1, A = 1). \quad (11)$$

Given  $\mu_{s_n}$ ,  $\mu_{s_m}$  is identified from  $(0, 0, 1)$ :

$$\mu_{s_m} = [(s_n + s_m)E(X|W = 0, M = 0, A = 1) - s_n\mu_{s_n}]/s_m. \quad (12)$$

Given  $\mu_{s_m}$  and  $\mu_{a_t}$ ,  $\mu_r$  is identified from  $(1, 1, 1)$ :

$$\mu_r = \frac{(r + s_m + a_t)E(X|W = 1, M = 1, A = 1) - s_m\mu_{s_m} - a_t\mu_{a_t}}{r}. \quad (13)$$

QED

There is one overidentifying condition for each characteristic at this stage. Propositions 1 and 2 then imply that the parameters ( $w, r, s_n, s_m, l, a_t, \mu_r, \mu_{s_n}, \mu_{s_m}, \mu_l, \mu_{a_t}$ ) are identified. We can thus study the effectiveness of magnet programs to attract and retain students. Moreover, the fraction of households that are at risk is the key parameter that measures the selective attrition between lottery winners and losers. Analyzing at-risk households is also important for the district and policy makers. Many urban districts have struggled in the past to retain students from higher-socioeconomic backgrounds. Magnet programs are perceived to be one possible solution to this problem. It is therefore important to quantify the impact of magnet programs on household retention in the district.

### C. Identification of Treatment Effects

We now turn to the analysis of identification of causal treatment effects of magnet programs on educational and behavioral outcomes. We assume that the researcher observes outcomes,  $T$ , only for students who remain in the school district; that is, we do not observe outcomes for leavers and at-risk households that lose the lottery. As a consequence, we face a missing-data problem in the analysis.

It is useful to assume initially that we observe the latent household type. Table 2 provides a summary of the relevant conditional expectations.<sup>15</sup> Conditioning on lottery outcomes, there are 10 conditional expectations. Three of these pertain to outcomes that are not observed since students in these latent groups leave the school district ( $T_2$ ). The remaining seven conditional expectations relate to household types that remain in the district.

From table 2, it is evident that even if we observed the latent types, there is little hope in identifying  $ATE_{S_n}$ ,  $ATE_R$ ,  $ATE_L$ , or  $ATE_{A_t}$ . For stayers who never attend the magnet program, we cannot identify  $E[T_1|S_n = 1]$ . For students at risk, we cannot identify  $E[T_0|R = 1]$ . For leavers, we cannot identify either  $E[T_1|L = 1]$  or  $E[T_0|L = 1]$ . For always-takers, we never observe  $E[T_0|A_t = 1]$ . Without imposing additional assumptions on the selection of students into latent groups,  $ATE_{S_n}$ ,  $ATE_R$ ,  $ATE_L$ , and  $ATE_{A_t}$  are not identified. Attention, therefore, focuses on identification of  $ATE_{S_m}$ .

This treatment effect is of interest to policy makers since the complying stayers account for the majority of students who attend magnet schools at any point in time. Our estimates suggest that 60%–70% of all the students in our sample of applicants to magnet programs and approximately 70%–80% of all attending students fall into that category. The school district and

<sup>15</sup> Note that we are implicitly assuming that the mean performance of stayers who would decline lottery admission is the same whether they win or lose the lottery, i.e.,

$$E[T_0|S_n = 1, W = 1] = E[T_0|S_n = 1, W = 0] = E[T_0|S_n = 1].$$

**Table 2**  
**Mean Outcomes Conditional on Type**

	Complying Stayers	Noncomplying Stayers	At Risk	Leavers	Always- Takers
$W = 1$	$E[T_1 S_m = 1]$	$E[T_0 S_n = 1]$	$E[T_1 R = 1]$	$E[T_2 L = 1]$	$E[T_1 A_t = 1]$
$W = 0$	$E[T_0 S_m = 1]$	$E[T_0 S_n = 1]$	$E[T_2 R = 1]$	$E[T_2 L = 1]$	$E[T_1 A_t = 1]$

NOTE.— $T_2$  is never observed.

policy makers are obviously interested in finding out whether the magnet programs improve outcomes for the majority of students who are attending the program.

Note that  $ATE_{S_m}$  would be identified if types were not latent. Of course, household types are not observed, and as a consequence, identification of  $ATE_{S_m}$  is not straightforward. One key result of this article is that the local average treatment effect for compliers is not point-identified if there is selective attrition.

Consider the case in which there is selective attrition ( $r \neq 0$ ). We only observe mean outcomes for the students conditional on  $W$ ,  $M$ , and  $A$ . For students who win the lottery and attend the magnet school, we observe

$$\begin{aligned} E[T|W = 1, M = 1, A = 1] \\ = \frac{s_m E[T_1|S_m = 1] + r E[T_1|R = 1] + a_t E[T_1|A_t = 1]}{s_m + r + a_t}. \end{aligned} \quad (14)$$

For students who lose the lottery and attend the magnet school, we observe

$$E[T|W = 0, M = 1, A = 1] = E[T_1|A_t = 1]. \quad (15)$$

We also observe mean performance of stayers who lose the lottery:

$$\begin{aligned} E[T|W = 0, M = 0, A = 1] \\ = \frac{s_m E[T_0|S_m = 1] + s_n E[T_0|S_n = 1]}{s_m + s_n}. \end{aligned} \quad (16)$$

Finally, we also observe the mean performance of stayers who win the lottery and decline to enroll in the magnet program:

$$E[T|W = 1, M = 0, A = 1] = E[T_0|S_n = 1]. \quad (17)$$

Equations (16) and (17) imply that we can identify  $E[T_0|S_m = 1]$  and  $E[T_0|S_n = 1]$  since  $s_n$  and  $s_m$  have been identified before. Equation (15) implies that we can identify  $E[T_1|A_t = 1]$ . However, equation (14) then implies that we cannot separately identify  $E[T_1|S_m = 1]$  and  $E[T_1|R = 1]$ .

This result illustrates that attrition per se is not the problem. If the fraction of at-risk households is negligible (i.e.,  $r = 0$ ), identification is achieved even if the fraction of leavers is large.<sup>16</sup> The lack of point identification arises from the at-risk households, which cause the selective attrition problem. Selective attrition is a problem only if at-risk households have different mean outcomes than compliers.<sup>17</sup>

Since point identification is no longer feasible when selective attrition is not negligible, attention focuses on set identification and the construction of bounds.

PROPOSITION 3.

- i. Suppose we have an upper bound, denoted by  $T_1^u$ , for  $E[T_1|R = 1]$ ; that is,  $T_1^u$  satisfies  $E[T_1|R = 1] \leq T_1^u$ . We can then construct a lower bound for  $E[T_1|S_m = 1]$  and  $ATE_{S_m}$ .
- ii. Suppose we have a lower bound, denoted by  $T_1^l$ , for  $E[T_1|R = 1]$ ; that is,  $T_1^l$  satisfies  $E[T_1|R = 1] \geq T_1^l$ . We can then construct an upper bound for  $E[T_1|S_m = 1]$  and  $ATE_{S_m}$ .

*Proof.* Consider the first part of the statement. Equation (14) then implies that

$$\begin{aligned}
 E[T_1|S_m = 1] &= \frac{s_m + r + a_t}{s_m} E[T|W = 1, M = 1, A = 1] \\
 &\quad - \frac{rE[T_1|R = 1] + a_tE[T_1|A_t = 1]}{s_m} \\
 &\geq \frac{s_m + r + a_t}{s_m} E[T|W = 1, M = 1, A = 1] \\
 &\quad - \frac{rT_1^u + a_tE[T_1|A_t = 1]}{s_m},
 \end{aligned} \tag{18}$$

where the last inequality follows from  $E[T_1|R = 1] \leq T_1^u$ . Since all terms in the last inequality are identified, we conclude that we can construct a lower bound. We therefore define our lower bound as the parameter value for which equation (18) holds with equality.

<sup>16</sup> Recall that if  $r = l = 0$ , our research design simplifies to the one considered in Angrist et al. (1996).

<sup>17</sup> We can generalize this result by assuming that  $E[T_1|S_m = 1, X] \neq E[T_1|R = 1, X]$ , i.e., by conditioning on some observables  $X$ . If controlling for selection on observables is sufficient to deal with the selection problem, a matching approach can be justified. For a discussion of matching estimators, see, among others, Rosenbaum and Rubin (1983), Heckman, Ichimura, and Todd (1997), and Abadie and Imbens (2006).

Replacing  $T_1^u$  with  $T_1^l$  and reversing the inequality yields the upper bound. QED

We thus obtain a lower and an upper bound for  $E[T_1|S_m = 1]$ . These bounds are sharp in the sense that without additional information we cannot improve on them. It is easy to see that for any value between the lower and upper bounds, there is a data-generating process such that  $T_1^l \leq E[T_1|R = 1] \leq T_1^u$  that is consistent with the observed means in equation (18). This follows essentially from the fact that equation (18) is linear in the observed means.

There are different ways of constructing lower and upper bounds depending on the outcome variable. A plausible assumption for the construction of an upper bound of the mean treatment effect is that the at-risk households are at least as good as the compliers,  $T_1^l = E[T_1|S_m = 1] \leq E[T_1|R = 1]$ . A better approach that we explore in this article is to bound outcomes using known percentiles of the outcome distribution. These types of aggregate distributions are often available in applications in education at the state level, as we discuss in detail in the next section.

Alternatively, we use a trimming approach as suggested by Lee (2009). This approach is applied in our context by first ordering magnet students from lowest to highest performance on the outcome variable being studied. Then treatment observations are dropped from the sample on the basis of both the proportions of missing data in the control and treatment groups and the distribution of the outcome variable being bounded.

We have seen that selective attrition implies that we have to focus on the construction of bounds since point identification is not feasible. It is therefore important to have a simple test to determine whether  $r$  is zero. If  $r = 0$ , treatment effects are point-identified and can be estimated using standard linear instrumental variable (IV) estimators. A simple way to estimate  $r$  is to regress  $A_i$  on  $W_i$ . The slope coefficient in that regression is equal to  $r$ . At minimum, researchers who work with lottery data in educational applications should run this regression and test whether one of the key identifying assumptions of the IV estimator is valid. If we reject the null that  $r$  is equal to zero, the bounds analysis suggested in this article is more appropriate than IV estimation.

#### D. A Generalized Method of Moments (GMM) Estimator

Let  $\theta$  denote the parameter vector that includes the fraction of the latent types, the means of the characteristics, and the lower and upper bounds of the various treatment effects. Suppose that we observe a random sample of  $N$  applicants to an education program, indexed by  $i$ . We view these as  $N$  independent draws from the underlying population of all applicants to this program. Let  $W_i$ ,  $M_i$ ,  $A_i$ , and  $X_i$  now denote the random variables that correspond to observation  $i$ . The proofs of identification are constructive.

Replacing population means by sample means thus yields consistent estimators for the parameters of interest. It is useful to place the estimation problem within a well-defined GMM framework. This allows us to estimate simultaneously all parameters and compute asymptotic standard errors. We can estimate the fractions of each latent type on the basis of moment conditions derived from the choice probabilities in table 1. Define

$$f_1(A_i, M_i, W_i|\theta) = \begin{cases} W_i M_i A_i - \omega(r + s_m + a_t) \\ W_i(1 - M_i)A_i - \omega s_n \\ W_i(1 - M_i)(1 - A_i) - \omega \\ (1 - W_i)M_i A_i - (1 - \omega)a_t \\ (1 - W_i)(1 - M_i)A_i - (1 - \omega)(s_n + s_m), \end{cases}$$

and note that

$$E \left[ \frac{1}{N} \sum_{i=1}^N f_1(A_i, M_i, W_i|\theta_0) \right] = 0,$$

where  $\theta_0$  denotes the true parameter value. Similarly, we can estimate the mean characteristics of each type. Define

$$f_2(A_i, M_i, W_i, X_i|\theta) = \begin{cases} W_i M_i A_i X_i - \omega[r\mu_r + s_m\mu_{s_m} + a_t\mu_{a_t}] \\ W_i(1 - M_i)A_i X_i - \omega s_n\mu_{s_n} \\ W_i(1 - M_i)(1 - A_i)X_i - \omega l\mu_l \\ (1 - W_i)M_i A_i X_i - (1 - \omega)a_t\mu_{a_t} \\ (1 - W_i)(1 - M_i)A_i X_i - (1 - \omega)[s_n\mu_{s_n} + s_m\mu_{s_m}] \\ (1 - W_i)(1 - M_i)(1 - A_i)X_i - (1 - \omega)[r\mu_r + l\mu_l], \end{cases}$$

and note that

$$E \left[ \frac{1}{N} \sum_{i=1}^N f_2(A_i, M_i, W_i, X_i|\theta_0) \right] = 0.$$

Finally, we can construct additional orthogonality conditions to construct both upper and lower bounds. Consider first the case of estimating an upper bound for compliers, denoted by  $E[T_1^u|S_m = 1]$ , by setting the

lower bound for treatment for at-risk  $E[T_1|R = 1]$  types equal to compliers ( $E[T_1|S_m = 1]$ ), denoted by  $T_1^l$ . Define

$$f_3(A_i, M_i, W_i, T_i|\theta) = \begin{cases} T_i W_i M_i A_i - w(s_m E[T_1^u|S_m = 1] + r T_1^l \\ \quad + a_t E[T_1|A_t = 1]) \\ T_i(1 - W_i)M_i A_i - (1 - w)a_t E[T_1|A_t = 1] \\ T_i(1 - W_i)(1 - M_i)A_i - (1 - w)(s_m E[T_0|S_m = 1] \\ \quad + s_n E[T_0|S_n = 1]) \\ T_i W_i(1 - M_i)A_i - w s_n E[T_0|S_n = 1], \end{cases}$$

and we have

$$E\left[\frac{1}{N} \sum_{i=1}^N f_3(A_i, M_i, W_i, T_i|\theta_0)\right] = 0.$$

Similarly, we can construct an orthogonality condition for the lower bound if we use the 95th percentile outcome for  $T_1^u$ . This value comes from state-level data for test scores and from our sample of nonmissing data at the district level for all other outcomes. Combining all orthogonality conditions, we can estimate the parameters of the model using a GMM estimator (Hansen 1982).

The main advantage of the GMM framework is that we can estimate all parameters jointly by imposing all relevant orthogonality conditions. Moreover, it is straightforward to obtain standard errors for the upper and lower bounds using a GMM framework. For each household characteristic we add, we obtain one overidentifying restriction. We implement our GMM estimator using a standard two-step approach, where the optimal weighting matrix is estimated on the basis of a consistent first-step estimator. Standard errors can be computed using the standard result for optimally weighted GMM estimators.<sup>18</sup>

There are different approaches to construct confidence intervals for partially identified models. One approach constructs a confidence set that includes each element in the identified set with fixed probability. The second approach constructs a confidence set that contains the entire identified

<sup>18</sup> Many of the parameters of the model—especially all parameters that characterize the fraction of latent types—can be estimated using linear estimators. An appendix is available on request that shows exactly how to set up the linear estimators.

set with fixed probability.<sup>19</sup> In our application, the identified set can be described by a set of closed intervals. As a consequence, we report only point estimates for the end points of these intervals and estimate standard errors for these estimates.

Thus far we have considered the problem of estimating causal effects using data from one lottery. In practice, researchers often need to pool data from multiple lotteries to obtain large enough sample sizes. We discuss in detail in appendix A of this article the problems that are encountered when aggregating across lotteries. Using a suitable weighting procedure, we show that we can estimate weighted averages of the underlying parameters of the model. Weights can be chosen in accordance with the objectives of the policy or decision maker.

#### IV. Data

We focus on magnet programs that are operated by a mid-sized urban school district that prefers not to be identified. Magnet schools emerged in the United States in the 1960s. Magnet schools are designed to draw students from across normal attendance zones. In contrast, a feeder school typically admits only students who live inside the attendance zone. As a consequence, the composition of feeder schools reflects residential choices of parents and is largely driven by the composition of local neighborhoods.

Magnet schools were initially used as a way to reduce racial segregation in public schools. More recently, magnet programs have been viewed as attractive options to increase school choice, to retain students with higher socioeconomic backgrounds in public schools, and to increase student achievement. In some cases, magnet programs are housed in separate schools. But they can also be a program within a more comprehensive school. Magnet programs offer specialized courses or curricula. There are magnet programs for all grade levels in our district. We consider only magnet programs that are academically oriented. These magnet programs typically provide specialized education in mathematics, the sciences, languages, or humanities. Other magnet programs have a broader focus on topics such as international studies or performing arts.

Every academic year, interested students submit applications for one magnet program of their choice. Some magnet programs in the district have a competitive entrance process, requiring an entrance examination, interview, or audition. We do not include these magnet programs in this study since the admission procedure does not use randomization. Instead we focus on magnet programs that do not have competitive entrance procedures. If the number of applications submitted during registration for any magnet pro-

<sup>19</sup> For a more detailed discussion, see, e.g., Imbens and Manski (2004) and Chernozhukov, Hong, and Tamer (2007).



gram exceeds the number of available spaces, the district holds a lottery to determine the order in which applicants will be accepted.

Many of the magnet schools in our district are vocational in nature. These programs are always undersubscribed and are not included in our study. Nearly every academically oriented magnet school held at least one binding lottery over the course of the study and is included in our sample.

In the case of oversubscription, a computerized random selection determines each student's lottery number. The lottery is binding in the sense that students with lower numbers are accepted and higher-numbered students are rejected. There is a clear cutoff number that separates the groups.<sup>20</sup> We do not observe students attending magnet schools who lose the lottery; that is, there are no always-takers in our sample.

To preserve racial balance in the magnet programs, separate lotteries are held for black students and other students. Some programs also have preferences for students with siblings already attending the magnet programs or for students who live close to the school. Separate lotteries are held for those students with an acceptable preference category for each magnet program. All in all, each lottery is held for a given program, in a given academic year, separately by race, and, finally, separately by preference code.

Lottery winners (lotteried-in) have the option to participate in the magnet program, with the ultimate choice of participation left to the student and his family. Lottery losers (lotteried-out) do not have this option and thus must make their schooling choice without the availability of the magnet option. With a fair and balanced lottery, the winners and losers will be determined by chance, thus creating two groups that are similar to each other on both observable and unobservable characteristics.

The district granted us access to its longitudinal student database. We use data from the 1999–2000 school year through 2005–6. In addition to demographic data, the database contains detailed information about educational outcomes. This information is linked to each student by a unique identification number. The demographic characteristics for the students include race, gender, free/reduced-lunch eligibility, and addresses. To be eligible for free lunch, households must have income below 130% of the poverty line. Reduced-lunch eligibility requires income below 185% of the poverty line.<sup>21</sup> Using the addresses, we can assign census tract-level variables to each student. We use two community characteristics that measure the socioeconomic composition of the neighborhoods in which students

<sup>20</sup> Strictly speaking, the win probabilities depend on the ordering of students on the wait list. However, these effects are probably small. As a consequence, the literature ignores these issues.

<sup>21</sup> The race variable is one if a student is African American and zero otherwise. The gender variable is one for girls and zero for boys.

reside. Poverty is the percentage of adults in the student's census tract with an income level below the poverty line. Education is the percentage of adults in the student's census tract with at least a college degree.

As pertaining to student educational outcomes, the database includes the school of attendance in each year and standardized scores for the state assessment tests. In addition, we observe a variety of behavioral outcome measures such as offenses, suspensions, and absences. The district has a code of student conduct that classifies two types of offenses, conveniently labeled level 1 infractions and level 2 infractions. Level 1 infractions are those of a less serious nature that do not necessarily pose a threat to the health, safety, or property of any person. These include truancy and class cuts, minor class disruption, teasing, refusal to participate in class, refusal to comply with staff directives, inappropriate language, and littering. Staff handle and correct level 1 offenses on their own without informing higher-level administrators. Level 2 infractions are those of a serious nature that may pose a threat to the health, safety, or property of any person. These include disruption of school, damage of school property, assault of a school employee or another student, weapons or drug possession, sexual harassment, academic dishonesty, bullying, and fighting. Staff are required to notify an administrator when a level 2 offense occurs. The administrator is then charged with the completion of an investigation and subsequent determination of consequences for the offender. Disciplinary action can include in-school suspensions, out-of-school suspensions, alternative education placement, and expulsion. These requirements hold for every school in the district at all levels, regardless of magnet designation or not.

In our data, the number of suspension days due to each offense is listed, and a very small minority reveal zero suspension days for the incident. Over 60% of the observations show 1 suspension day and 96% show 3 days or less of suspension. There is not any explicit description of the infraction, nor is there any clarification as to whether the offenses are level 1 or level 2. However, in light of the writing and required notification policy explained in the code of student conduct and the fact that nearly all of these incidents result in at least 1 day of suspension, we believe that the events we call offenses in the data files are level 2 infractions. In other words, these are highly disruptive and definitely nonconducive to the educational environment. They are relatively extreme behavioral problems that would be properly identified, in the same way, in every school.

The database also contains the outcomes of the magnet lotteries. We do not observe test scores or behavioral outcome measures for students outside of the district. Table 3 shows descriptive statistics for the entire sample used in this study as well as three important subsamples that we also consider in estimation.<sup>22</sup> We consider only binding lotteries in this

<sup>22</sup> We imputed absences and tardies for a small sample of students. Also note that behavioral variables are not observed for students who apply from nonpublic

**Table 3**  
**Descriptive Statistics**

Variable	Entire Sample ( <i>N</i> = 2,054)	Elementary School ( <i>N</i> = 820)	Middle School ( <i>N</i> = 457)	High School ( <i>N</i> = 777)
Gender	.51 (.50)	.51 (.50)	.51 (.50)	.51 (.50)
Race	.75 (.44)	.59 (.49)	.79 (.40)	.88 (.32)
Subsidized lunch	.33 (.47)	.33 (.47)	.35 (.48)	.32 (.47)
Poverty	.23 (.14)	.22 (.14)	.23 (.14)	.24 (.15)
Education	.29 (.19)	.34 (.22)	.28 (.18)	.25 (.14)
Offenses	.99 (2.23)	.18 (.99)	1.15 (2.32)	1.67 (2.71)
Suspension days	1.88 (4.71)	.29 (1.62)	1.97 (4.39)	3.32 (6.17)
Absences	13.28 (14.56)	8.74 (7.96)	10.30 (8.54)	19.30 (19.30)
Tardies	7.31 (13.10)	3.94 (7.03)	8.66 (12.89)	9.70 (16.55)
Win percentage	61.8	52.1	53.2	77.1

NOTE.—Standard deviations are reported in parentheses.

research. In total, over the time frame of the data, there are 173 binding lotteries with 1,269 students lotteried-in and 785 students lotteried-out.

Before we implement the estimators, we check whether the lotteries are balanced on student observables. While assignment within lotteries may be random, participation in a lottery is not. To make use of within-lottery randomness and not the between-lottery nonrandomness, we perform a check for balance by running a lottery fixed-effect regression for each observable characteristic as a dependent variable with acceptance as the only independent variable other than the fixed effects. Separate lotteries are held by race, so race is left out of the balance analysis. We test every other observable student characteristic in the data set.

Following Cullen et al. (2006), we use equation (19) to determine whether the lottery is balanced:

$$X_i = \beta_1 W_i + \sum_{j=1}^J I_{ij} \beta_{2j} + v_i, \quad (19)$$

schools. Thus the means of the last four variables (excluding win percentage) in table 3 reflect means of magnet applicants in district public schools at the time of application. The values for all applicants are used for the means of the other demographic characteristics.

where  $X_i$  is the observable characteristic of interest,  $W_i$  is a dummy equal to one if student  $i$  wins lottery  $j$ ,  $I_{ij}$  is an indicator variable equal to one if student  $i$  participated in lottery  $j$ , and  $v_i$  is the error term. We estimate a separate regression for each observable. The coefficient  $\beta_1$  determines the fairness of the lottery system. If we cannot reject the null hypothesis that it is equal to zero, then acceptance into a magnet is not determined by the value of that particular student observable,  $X_i$ .

The first column of table 4 shows the results when all students in all binding lotteries are included in the regressions. The coefficient  $\beta_1$  is not significant for any tested variable at 10%. The second through fourth columns contain the other subsamples of interest. We find that the estimates of  $\beta_1$  are not significantly different from zero.

In addition to the tests reported in table 4, we have also implemented joint tests using seemingly unrelated regressions. The  $p$ -values of the corresponding  $F$ -tests are .933 for the full sample, .704 for the elementary schools, .816 for the middle schools, and .522 for the high school subsample. We thus conclude that the joint tests fail to reject the null hypothesis that all coefficients are zero. We thus find that the lotteries are fair, creating separate winner and loser groups that are similar in observed characteristics. Any differences between winners and losers are small and statistically insignificant. This holds for the overall population in binding lotteries and for the smaller subsamples that were tested.

The design of the preferences codes in the admission process implies that there is no variation of race within lotteries. The lottery fixed effects will capture the effects of race. We therefore cannot conduct the standard balance analysis for race. To get some additional insights into differences among racial groups, we computed win percentages by race and report them in table 5.

The district is approximately 55% black. In addition, nonblack students live in areas with better neighborhood schools, even within the same dis-

**Table 4**  
**Lottery Balance Result**

Variable	Entire Sample	Elementary School	Middle School	High School
Gender	.0053 (.0262)	.0366 (.0384)	-.0183 (.0559)	-.0257 (.0469)
Subsidized lunch	.0056 (.0229)	.0111 (.0322)	-.0501 (.0482)	.0385 (.0431)
Poverty	-.0050 (.0068)	-.0023 (.0092)	.0044 (.0136)	-.0160 (.0135)
Education	.0041 (.0078)	.0110 (.0127)	-.0038 (.0165)	-.0007 (.0125)

NOTE.—Estimated standard errors are reported in parentheses.

**Table 5**  
**Lotteries by Race**

Level	Black Wins	Black Losses	Other Wins	Other Losses	Win Percentage	
					Black	Other
Elementary	233	252	194	141	48.04	57.91
Middle	163	200	80	14	44.90	85.11
High	530	155	69	23	77.37	75.00
All	926	607	343	178	60.40	65.83

trict, somewhat mitigating their interest in magnet schools. As a consequence, there are many more black applicants in our study. Black applicants have lower overall win percentages.

## V. Empirical Results

### A. Attraction, Retention, and Selective Attrition

To study the importance of selective attrition in our sample, we implement a number of different estimators. First, we use a GMM estimator that imposes only the orthogonality conditions that identify the fraction of latent household types. Then we add the orthogonality conditions that can be used to estimate the mean characteristics of the types. The characteristics include race, gender, free or reduced lunch, poverty, and college education. Recall that the last two measures are based on neighborhood characteristics as reported by the US Census. We report estimates for three samples, which include all students who applied to an oversubscribed magnet program that is associated with an elementary school, middle school, and high school, respectively. We pool across all lotteries in each sample and therefore use the weighted estimator discussed in appendix A. Tables 6 and 7 report the point estimates and estimated standard errors for each of the three samples.

Comparing the estimates in the upper and lower panels of table 6 clearly allows us to evaluate whether there are efficiency gains that arise when using a GMM estimator.<sup>23</sup> We find that there are significant efficiency gains in the estimates of two key parameters, the fraction of compliers and the fraction at risk. Estimated standard errors are up to 50% larger when one ignores the additional orthogonality conditions. Our framework does not generate many overidentifying restrictions. For each household characteristic we add, we obtain one overidentifying restriction. This is the case be-

<sup>23</sup> This comparison is also interesting since the GMM estimates and associated standard errors in the upper panel are identical to the results that could be obtained using simpler linear estimators.

**Table 6**  
**Empirical Results: Selective Attrition**

	Fraction At Risk	Fraction Stay, Attend	Fraction Stay, Non	Fraction Leave
First set of orthogonality conditions:				
Elementary	.25 (.04)	.61 (.05)	.06 (.01)	.08 (.01)
Middle	.12 (.15)	.60 (.16)	.24 (.04)	.04 (.01)
High	.15 (.09)	.70 (.09)	.08 (.01)	.06 (.01)
First and second set of orthogonality conditions:				
Elementary	.25 (.04)	.61 (.04)	.06 (.01)	.08 (.01)
Middle	.12 (.05)	.61 (.06)	.24 (.04)	.04 (.01)
High	.14 (.06)	.72 (.06)	.08 (.01)	.06 (.01)

NOTE.—Estimated standard errors are reported in parentheses.

**Table 7**  
**Empirical Results: Characteristics**

	At Risk	Stay, Attend	Stay, Non	Leave
Gender:				
Elementary	.57 (.09)	.47 (.03)	.55 (.11)	.47 (.08)
Middle	.85 (.34)	.43 (.06)	.50 (.08)	.31 (.13)
High	.55 (.34)	.57 (.05)	.49 (.08)	.41 (.08)
Race:				
Elementary	.50 (.09)	.70 (.04)	.39 (.11)	.18 (.07)
Middle	.99 (.41)	.80 (.05)	.80 (.06)	.28 (.14)
High	.89 (.41)	.93 (.03)	.85 (.07)	.79 (.06)
Free or reduced lunch:				
Elementary	.12 (.04)	.43 (.03)	.19 (.07)	.07 (.04)
Middle	.26 (.15)	.47 (.06)	.26 (.09)	.07 (.06)
High	.15 (.11)	.39 (.04)	.25 (.06)	.12 (.05)
Poverty:				
Elementary	.21 (.03)	.23 (.01)	.20 (.04)	.14 (.01)
Middle	.24 (.10)	.24 (.02)	.23 (.02)	.13 (.02)
High	.28 (.12)	.25 (.01)	.24 (.02)	.19 (.02)
Education:				
Elementary	.40 (.05)	.29 (.02)	.41 (.05)	.53 (.04)
Middle	.20 (.11)	.29 (.02)	.30 (.03)	.55 (.08)
High	.27 (.14)	.25 (.01)	.21 (.02)	.36 (.03)

NOTE.—Estimated standard errors are reported in parentheses.

cause we observe the conditional expectation of  $X$  for six observed types. But these conditional expectations are functions of the means of the five latent types. Hence, we have little reason to believe that we are suffering from small-sample problems that can arise when the number of orthogonality conditions is too large. We thus conclude that our approach of jointly estimating the model using GMM is preferable to simpler methods.

Table 6 reveals some interesting new insights into the importance of selective attrition in our application. Recall that the fraction of households at risk is the key parameter that captures selective attrition. We find that selective attrition is substantial and ranges between 12% and 25% across our three samples. We also find that the majority of students will stay in the district regardless of the outcome of the lottery. The majority, 61%–71%, will attend the magnet program if they win the lottery. The fraction of households that will leave the district regardless of the outcome of the lottery ranges between 4% and 8%. Overall, these results suggest that most households consider the magnet programs desirable. We conclude that magnet programs are effective tools for attracting and retaining households and students.

Equally interesting are the observed mean characteristics of the latent types of households reported in table 7. These and the ones reported in the lower part of table 6 are the results from the first and second set of orthogonality conditions ( $f_1$  and  $f_2$ ). For each characteristic, the differences across household types (at risk, leavers, stayers) are statistically significant. We find that at-risk households are, on average, less likely to be African American and to be on free- or reduced-lunch programs than households that are stayers. Moreover, they come from better-educated neighborhoods.<sup>24</sup> These differences are more pronounced at the elementary school level, where the fraction of at-risk households is the greatest. We thus conclude that magnet programs are effective devices for the school district to retain more affluent households. Not surprisingly, the leavers are the most affluent group and come from neighborhoods with the highest levels of education. These households may just apply to the magnet programs as a backup option in case their students should unexpectedly not be admitted to an independent, charter, or parochial school.<sup>25</sup>

The demographic differences, summarized above, between at-risk students and stayers drive our assumptions on the bounds. Poor minority students are known to perform poorly in school compared to wealthier majority peers (Dobbie and Fryer 2009). Therefore, our upper-bound estimation assumes that the performance of at-risk students is only as good as that of the stayers, while the lower-bound estimation assumes that the at-risk students are in the 95th percentile of the outcome distribution.

Tables 3 and 7 permit interesting comparisons across grade levels. From table 3, we see that elementary and middle school lotteries are somewhat more competitive than high school lotteries. The former have average win

<sup>24</sup> Note that the differences in household characteristics are statistically significant from zero at all conventional levels.

<sup>25</sup> It could also be that these households left the district because of job transfers or other issues unrelated to schools.

rates of 52% and 53%, respectively, while the latter have an average win rate of 77%. Table 7 provides information about types by grade levels. We see that elementary programs attract a clientele from more highly educated neighborhoods. The fraction of African American families is also lower among applicants to elementary school lotteries. Not surprisingly, we find that the fraction of at-risk families and the fraction of leavers are also higher among elementary school students. These findings highlight the fact that, among the magnet school applicants, the market for elementary school education is more competitive than the market for high school education.

### B. Treatment Effects

We have seen in the previous section that the fraction of at-risk households is large and significantly different from zero in our application in all three samples. Moreover, households that are at risk of leaving the district have more favorable socioeconomic characteristics than other types except for leavers. As a consequence, we conclude that selective attrition cannot be ignored in this application. Since treatment effects are set-identified only when selective attrition matters, we implement our bounds estimators. For comparison purposes, we also report the IV estimates that ignore selective attrition.

We start our analysis by focusing on achievement effects. The main problem encountered in this part of the analysis arises as a result of missing data. This is largely the case because standardized achievement tests were conducted only in grades 5, 8, and 11 during most of our sample period. For our middle school sample, there are only 155 observations for which we have prior test scores. For the high school sample, the reduction is of similar magnitude.<sup>26</sup> Ultimately, we have test score outcomes for 213 middle school students (eighth-grade exam) and 203 high school students (eleventh-grade exam). Table 8 summarizes our main findings using standardized test scores in reading and mathematics as outcome variables.

We find that the point estimates of the upper and lower bounds point to positive treatment effects, but sample sizes are too small to provide precise estimates. While few researchers would advocate the use of the simple IV estimator in the presence of selective attrition, it is useful to compare the results of our bounds analysis with the IV approach. One surprising finding is that the simple IV estimates suggest statistically significant positive treatment effects. Our bounds analysis reveals that this inference is not correct.

<sup>26</sup> Moreover, we find some evidence that lower-performing students are more likely to drop out of the sample, perhaps because they drop out of school.



**Table 8**  
**Empirical Results: Achievement**

	ATE <sub><math>s_m</math></sub>		
	Upper Bound	Lower Bound	IV
Reading:			
Middle	66.25 (118.30)	3.68 (172.69)	139.71 (77.33)
High	77.05 (64.79)	-25.09 (136.24)	81.97 (47.17)
Mathematics:			
Middle	180.89 (124.89)	91.08 (183.69)	138.56 (63.63)
High	87.09 (57.62)	-24.22 (148.00)	94.30 (40.70)

NOTE.—Estimated standard errors are reported in parentheses.

We next turn our attention to behavioral outcomes measured a year after the lotteries were conducted.<sup>27</sup> The main advantage of studying these outcomes is that we do not face the data limitations that we encounter with test scores. Comprehensive records of four important behavioral measures are available: suspensions, offenses, absences, and tardies.

Table 9 summarizes our main findings. Note that a negative treatment effect is a reduction in undesirable behavior and thus a good outcome. For elementary students, we find that magnet programs significantly reduce offenses and suspensions. There are no measurable effects on tardies and absences. We find that there are few significant treatment effects at the middle school level. The estimates themselves suggest that middle school magnet programs have a negative effect on offenses, no effect on suspensions, and possibly an increase in absences and tardies. Again, however, these estimates at the middle school level are generally not significant. For the high school sample, we find strong evidence that the magnet schools reduce absences and tardies while having no significant effects on offenses or suspensions. Comparing the IV estimates with the bounds, we find that the IV estimates are often of similar magnitude to our upper-bound estimates and have smaller estimated standard errors than the bound estimates.

We can derive the asymptotic limit of the IV estimator under our model specification. Setting the fraction of always-takers equal to zero (as is true for this application), we obtain

$$\hat{\beta}^{\text{IV}} \rightarrow \frac{s_m + s_n + r}{s_m + r}(A - B),$$

<sup>27</sup> Previously, Cullen et al. (2006) and Imberman (2011) have studied behavioral outcomes when examining school choice programs.

**Table 9**  
**Empirical Results: Behavioral Outcomes**

	ATE <sub>S<sub>m</sub></sub>		
	Upper Bound	Lower Bound	IV
Offenses:			
Elementary	-.28 (.09)	-.26 (.09)	-.26 (.09)
Middle	-.62 (.36)	-.48 (.36)	-.66 (.35)
High	-.03 (.34)	.28 (.39)	.20 (.31)
Suspensions:			
Elementary	-.49 (.15)	-.45 (.15)	-.47 (.14)
Middle	-.22 (1.17)	.00 (1.18)	-.56 (.77)
High	-.47 (.87)	.14 (.93)	.03 (.75)
Absences:			
Elementary	-2.26 (.90)	.98 (1.24)	-1.70 (.77)
Middle	1.98 (1.60)	4.16 (2.02)	1.82 (1.36)
High	-8.64 (3.32)	-5.35 (3.60)	-7.77 (2.55)
Tardies:			
Elementary	-.95 (.73)	.52 (.87)	-.98 (.59)
Middle	3.04 (1.82)	4.97 (2.07)	2.32 (2.07)
High	-7.90 (2.78)	-6.61 (2.87)	-9.41 (2.45)

NOTE.—Estimated standard errors are reported in parentheses.

where

$$\begin{aligned}
 A &= \left[ \frac{r}{r + s_n + s_m} E[T_1|R] + \frac{s_m}{r + s_n + s_m} E[T_1|S_m] \right. \\
 &\quad \left. + \frac{s_n}{r + s_n + s_m} E[T_0|S_n] \right], \\
 B &= \left[ \frac{s_m}{s_m + s_n} E[T_0|S_m] + \frac{s_n}{s_m + s_n} E[T_0|S_n] \right].
 \end{aligned}$$

Now if  $r = 0$ , then

$$\hat{\beta}^{IV} \rightarrow E[T_1|S_m] - E[T_0|S_m].$$

If  $r \neq 0$ , then there is no reason to believe that the IV estimator will always be within the bounds provided in this article.

We thus conclude that our bounds analysis is informative and demonstrates that magnet programs offered by the district improve behavioral outcomes. In particular, we find that offenses are significantly lower for elementary school students, while high school students have significantly better attendance and timeliness records. It is also important to note that the 95th percentile of all behavioral outcomes is zero except for high school absences (where it is one). Thus our lower-bound estimates for nearly all behavioral outcomes are the most pessimistic possible since they attribute flawless behavior to all who leave the district.

### C. Sensitivity Analysis

First, we investigate whether our results are sensitive to the choice of the percentile used to construct the lower bound. We can use the 90th or 99th statewide percentile on math and reading exams instead of the 95th percentile. The results are reported in table 10. We find that the results are qualitatively the same. We cannot reject the null hypothesis that the treatment effect is zero.

Next we consider the behavioral outcomes. The lower-bound estimates for the behavioral outcomes are exactly the same whether the 90th, 95th, or 99th percentile of the districtwide measures is used for tardies, offenses, and suspensions. In all cases, at all levels, the at-risk students who lose and leave are assumed to have no instances of any of these outcomes at all three percentiles. The lower-bound estimates for absences remain largely the same under all three percentile choices even though the assumptions for the at-risk types who lose and leave change a bit, moving up to three for the high school 90th percentile.

We also conduct two additional robustness checks. First, we find that our results are similar when we drop the five overidentifying conditions and implement an exactly identified estimator. For example, from the middle school sample, our estimate with (without) overidentifying conditions for the upper bound on offenses is  $-0.62$  ( $-0.64$ ), on suspension days  $-0.22$  ( $-0.23$ ), on absences  $1.98$  ( $1.61$ ), and on tardies  $3.04$  ( $3.02$ ). We also implemented the overidentified estimator for nonoptimal weighting matrices such as the identity matrix and again found insignificant differences. We thus conclude that these small-sample problems are not a problem in our application. Second, we explore heterogeneity in treatment. For middle school, there are 40 lotteries. Sixteen have win rates estimated above or equal to 0.5 and 24 have win rates less than 0.5. We can split the sample and determine separate treatment effects. From the full sample, our estimate for the upper bound on offenses is  $-0.62$  with an associated standard error of 0.36. Using the subsample with low win rates, we obtain an estimate of

**Table 10**  
**Results for Different Quantiles**

	Percentile		
	90th	95th	99th
Middle school math	129.15 (177.11)	91.08 (183.69)	35.66 (198.69)
Middle school reading	37.52 (167.98)	3.68 (172.69)	-59.35 (188.40)
High school math	-8.37 (131.69)	-24.22 (148.00)	-53.31 (179.01)
High school reading	-9.68 (121.37)	-25.09 (136.24)	-53.10 (164.81)
Elementary school absent	.58 (1.19)	.98 (1.24)	.98 (1.24)
Middle school absent	3.97 (1.97)	4.16 (2.02)	4.16 (2.02)
High school absent	-5.92 (3.50)	-5.54 (3.57)	-5.35 (3.60)

NOTE.—Estimated standard errors are reported in parentheses.

-0.49 (0.29). For high win rates, the estimate is -0.13 (0.25). The results are statistically similar for both subsamples. This result generally holds for all school levels and all outcomes.

Finally, we consider Lee's (2009) approach. Recall that one of the nice features of his estimator is that it does not require additional information. Instead it relies on trimming to construct an estimator for the lower and upper bounds of the treatment effect. It is therefore useful to implement this approach using the data from our application. Table 11 compares our estimates with those obtained from Lee's trimming method.<sup>28</sup> As we detail in appendix A, weighting is appropriate when estimating bounds using data from multiple lotteries. In implementing Lee's estimator, we do not weight lotteries by number of applicants.<sup>29</sup> Hence, the comparison in table 11 reflects both a difference in the approach to bounding and a difference in weighting, potentially confounding the two effects. For the outcomes considered in table 11, we have confirmed that the results from our weighted estimator are similar to those when we do not weight by lotter-

<sup>28</sup> The results are similar for other outcomes analyzed in this article. The four outcomes were chosen for the following reason. We have a large sample for elementary school offenses. Our point estimates suggest that the magnet schools may reduce offenses. For tardies, our estimates suggest no effect. The sample size for high school math is small, and our estimates suggest no significant treatment effect. Finally, the sample for middle school math is also small, but our estimates suggest that there may be a positive treatment effect.

<sup>29</sup> Lee's estimator has not yet been extended to estimate bounds when combining data from multiple lotteries, though it is surely possible to do so.

**Table 11**  
**Comparison with Lee's Estimator**

	Our Estimator	Lee's Estimator
Elementary offenses:		
Upper bound	-.28 (.09)	-.33 (.08) [347]
Lower bound	-.26 (.09)	-.27 (.08) [357]
Point estimate range	.02	.06
Simple TE CI	[-.46, -.08]	[-.49, -.11] $\hat{p} = .337$
Elementary school tardies:		
Upper bound	-.95 (.73)	-3.58 (.58) [217]
Lower bound	.52 (.87)	-.68 (.74) [306]
Point estimate range	1.47	2.90
Simple TE CI	[-2.38, 2.23]	[-4.72, .77] $\hat{p} = .362$
High school math:		
Upper bound	87.09 (57.62)	243.69 (301.97) [33]
Lower bound	-24.22 (148.00)	-150.83 (252.33) [33]
Point estimate range	111.31	394.52
Simple TE CI	[-314.30, 200.03]	[-645.40, 835.55] $\hat{p} = .660$
Middle school math:		
Upper bound	180.89 (124.89)	382.86 (286.89) [45]
Lower bound	91.08 (183.69)	65.11 (243.81) [48]
Point estimate range	89.81	317.75
Simple TE CI	[-268.95, 425.67]	[-412.76, 945.16] $\hat{p} = .426$

NOTE.—TE CI is the treatment effect confidence interval.

ies. This is not always the case, however. For example, for middle school reading, weighting by lotteries proves to be quite important.<sup>30</sup> Hence, it would be desirable in future work to extend the Lee estimator to weight lotteries. The two methods could then be compared on a common footing in applications with multiple lotteries.

Table 11 suggests that the empirical results are similar, but there is at least one noteworthy difference. We find that our estimator provides tighter bounds estimates for the magnet treatment effects than the one proposed in Lee (2009) in this application. Recall that our approach requires both additional data and additional assumptions to be valid. As a consequence, it is not that surprising that the approach proposed in this article sometimes yields tighter bounds.

Table 11 also reports the trimming proportions  $\hat{p}$  for Lee's estimator for all outcomes. Note that  $\hat{p}$  is the trimming proportion and is defined just as in Lee's paper. We find that the trimming rates are much greater in our application than in Lee's application, where  $\hat{p} = 0.068$ . The reason is

<sup>30</sup> Details are available on request.

that our proportion of nonmissing data between the control and treatment groups differs significantly since we never observe outcomes for those who leave the district. These students are exclusively contained in the control group since nobody can be in a magnet program yet outside of the district. The other main difference between our application and Lee's application is sample size. Lee reports over 3,000 observations in the treatment group before and after trimming. These samples are much larger than the ones in our application.

## VI. Conclusions

We have studied the effectiveness of magnet programs in a mid-sized urban district. Our empirical results suggest that selective attrition cannot be ignored in our application. We find that magnet programs are useful tools that help the district to attract and retain students from middle-class backgrounds. Finally, we have also studied the impact of magnet programs on achievement and a variety of behavioral outcomes. Our findings for achievement effects are mixed. While the point estimates of the bounds point to positive treatment effects, sample sizes are too small to provide precise estimates. For a variety of behavioral outcomes, we do not face these data limitations. Our evidence suggests that magnet programs often improve behavioral outcomes.

We acknowledge that our results may not be broadly applicable to all magnet programs. This is a general drawback of many policy papers, where a clean experimental design can ensure internal validity but lacks external validity. Nearly every academically oriented magnet school held at least one binding lottery over the course of the study. If magnet schools are better than regular feeder schools, students have strong incentives to get into any magnet. In that case, some students apply to lower-quality magnet schools in the hope that they face less stiff competition in the lottery process. This gives rise to selection across magnet programs that may affect peer qualities and other endogenous features of the school.

In some studies, it has been possible to link district-level data with state-level data. For example, it is possible to track students from the Boston school district, even if they leave the district, as long as they stay in Massachusetts. Moreover, researchers have been able to use statewide achievement tests as outcome measures. As a consequence, they have access to the same outcome measure for students who decided to stay in the district and who decided to leave. These data sets contain outcome measures for the at-risk types as well as the leavers if they stayed in the state. Unfortunately, the district that has provided us with the data is situated in a state that does not allow us to track students when they leave the district. That is quite common for other districts that have cooperated with researchers as well. Limited access to private school data is an even more pervasive problem.

As a consequence, we think that our framework applies to the vast majority of the potential applications. As better data-sharing arrangements become available, we will be able to expand our analysis to incorporate additional observables. Our framework permits estimation of the proportion of lottery applicants who remain in the district as a result of magnet admission. These students are denoted at risk in our framework. Hence, while we do not observe outcomes for students who leave the district, we are able to provide information about the retention effect of the magnet program. This is explicitly something that the district hopes the magnet programs achieve.

## Appendix A

### Aggregation

For a given magnet program, a separate lottery is conducted for each grade, and within a grade, separate lotteries may be conducted for different groups of applicants (e.g., by race). In such cases, sample sizes for individual lotteries may be relatively small, yielding lottery-specific estimates with low power. While outcomes for a particular lottery may be of interest, a district will typically be more concerned with evaluation at the program level rather than at the lottery level. Here we extend our analysis to permit investigation at the program level.

Suppose that there are  $j = 1, \dots, J$  lotteries governing access to a magnet program. A program may be a magnet school (or perhaps set of magnet schools) serving a particular range of school grades. Let  $w_j$  be the probability of winning lottery  $j$ , and, analogously to our previous notation, let  $a_{i,j}$ ,  $\ell_j$ ,  $r_j$ ,  $s_{m,j}$ , and  $s_{n,j}$  be the proportions of latent types in lottery  $j$ . Let  $N_j$  be the number of applicants to lottery  $j$  and  $N = \sum_j N_j$ . The share of lottery  $j$  is then  $n_j = N_j/N$ . When we extend our previous notation,  $W_{ij}$  equals one if applicant  $i$  to lottery  $j$  wins and zero otherwise,  $A_{ij}$  equals one if applicant  $i$  to lottery  $j$  attends a school in the district and zero otherwise, and  $M_{ij}$  equals one if applicant  $i$  to lottery  $j$  attends magnet school  $j$  and zero otherwise.

Let  $w = \sum_j n_j w_j$ ,  $a_t = \sum_j n_j a_{t,j}$ ,  $\ell = \sum_j n_j \ell_j$ ,  $s_m = \sum_j n_j s_{m,j}$ ,  $s_n = \sum_j n_j s_{n,j}$ , and  $r = \sum_j n_j r_j$ . Thus,  $w$ ,  $a_t$ ,  $\ell$ ,  $r$ ,  $s_m$ , and  $s_n$  are parameters denoting the share of each of the latent types at the program level. Our previous analysis applies to each lottery, establishing identification of  $a_{t,j}$ ,  $\ell_j$ ,  $r_j$ ,  $s_{m,j}$ , and  $s_{n,j}$  for all  $j$ . The  $n_j$  are known and nonrandom. Hence,  $w$ ,  $a_t$ ,  $\ell$ ,  $r$ ,  $s_m$ , and  $s_n$  are identified. We therefore focus on estimation and inference at the program level. Consider the following:

$$\frac{1}{N} \sum_{j=1}^J \sum_{i=1}^{N_j} \frac{W_{ij}(1 - M_{ij})A_{ij}}{w_j} \rightarrow \sum_{j=1}^J n_j s_{n,j} = s_n. \quad (\text{A1})$$

Proceeding analogously for other latent types, we obtain the orthogonality conditions below for estimating program-level parameters:

$$\begin{aligned}
& \frac{1}{N} \sum_{i=1}^N \frac{W_{i1}}{n_1} - w_1, \\
& \vdots \\
& \frac{1}{N} \sum_{i=1}^N \frac{W_{ij}}{n_j} - w_j, \\
& \frac{1}{N} \sum_{j=1}^J \sum_{i=1}^{N_j} \frac{W_{ij} M_{ij} A_{ij}}{w_j} - (r + s_m + a_t), \\
& \frac{1}{N} \sum_{j=1}^J \sum_{i=1}^{N_j} \frac{W_{ij} (1 - M_{ij}) A_{ij}}{w_j} - s_n, \\
& \frac{1}{N} \sum_{j=1}^J \sum_{i=1}^{N_j} \frac{W_{ij} (1 - M_{ij}) (1 - A_{ij})}{w_j} - l, \\
& \frac{1}{N} \sum_{j=1}^J \sum_{i=1}^{N_j} \frac{(1 - W_{ij}) M_{ij} A_{ij}}{1 - w_j} - a_t, \\
& \frac{1}{N} \sum_{j=1}^J \sum_{i=1}^{N_j} \frac{(1 - W_{ij}) (1 - M_{ij}) A_{ij}}{1 - w_j} - (s_n + s_m).
\end{aligned} \tag{A2}$$

Next, consider achievement. Let  $E[T_{1,j}|S_{m,j} = 1]$  denote the expected test score of a student who wins the lottery for program  $j$  and is a complying stayer. For simplicity let  $a_t = 0$ . Note that

$$\frac{1}{N_j} \sum_{i=1}^{N_j} T_i W_{ji} A_{ji} M_{ji} \rightarrow w_j \{r_j E[T_{1j}|R_j = 1] + s_{mj} E[T_{1j}|S_{m,j} = 1]\}. \tag{A3}$$

Using the same logic above and pooling over lotteries implies that

$$\frac{1}{N} \sum_{j=1}^J \frac{1}{w_j} \sum_{i=1}^{N_j} T_i W_{ji} A_{ji} M_{ji} \rightarrow \sum_{j=1}^J n_j \{r_j E[T_{1j}|R_j = 1] + s_{mj} E[T_{1j}|S_{m,j} = 1]\}. \tag{A4}$$

Now suppose that we have an upper bound  $U$  for  $E[T_{1j}|R_j = 1]$  for all  $j$ , that is,  $U \geq E[T_{1j}|R_j = 1]$  for all  $j$ . Hence



$$\sum_{j=1}^J n_j r_j E[T_{1j}|R_j = 1] \leq \sum_{j=1}^J n_j r_j U = rU. \quad (\text{A5})$$

Combining equations (A4) and (A5) and normalizing by  $s_m$ , we have

$$\frac{1}{s_m} \sum_{j=1}^J n_j s_{mj} E[T_{1j}|S_{m,j} = 1] \geq \frac{1}{s_m} \left[ \frac{1}{N} \sum_{j=1}^J \frac{1}{w_j} \sum_{i=1}^{N_j} T_i W_{ji} A_{ji} M_{ji} - rU \right]. \quad (\text{A6})$$

Hence we have constructed a lower bound for the weighted average of the treatment effect. Note that the weights depend on  $n_j$  and  $s_{mj}$ .

Using a lower bound  $L$  such that  $L \leq E[T_{1j}|R_j = 1]$  for all  $j$  yields an upper bound for the weighted treatment effect.

The weighting scheme we use is based on the win rates for each lottery. The win rates are themselves a function of the available seats and the number of applicants. This is similar to results in Frolich and Lechner (2010). The authors find aggregate average treatment effects separately by type by weighting each of 18 separate, local labor market type-specific outcomes with the number of participants, again by type, in each market. Our bounds for the complying stayers are in the same vein as Frolich and Lechner's estimates for the compliers.

## Appendix B

### Latent Types

Defining the potential outcomes  $A(W)$  and  $M(W)$ , the 16 possible types are given in table B1. Defiers are those who always go against their lotteried assignment; they attend the magnet when they lose the lottery, and they do not attend when they win.

It is not possible to attend a magnet school but not attend a district school. We can call this a feasibility assumption. This eliminates types 6, 7, 8, 10, and 12.

We then make the following assumptions to reduce the number of types to the five types used in the analysis.

#### ASSUMPTION 1.

- a. There are no individuals who would leave the district if they win the lottery but would stay if they lose, that is, no defiers with respect to moving.<sup>31</sup>
- b. There are no individuals who do not attend a nonmagnet district school if they lose the lottery but attend a nonmagnet district school if they win the lottery.
- c. There are no defiers (monotonicity assumption).

<sup>31</sup> This is the same monotonicity assumption used, e.g., in Lee (2009).

Table B1  
Potential Latent Types

1*	$A(0) = 1, M(0) = 0 \text{ and } A(1) = 1, M(1) = 0$	Never-taker ( $S_n$ )
2*	$A(0) = 1, M(0) = 1 \text{ and } A(1) = 1, M(1) = 1$	Always-taker ( $A$ )
3*	$A(0) = 1, M(0) = 0 \text{ and } A(1) = 1, M(1) = 1$	Complier ( $S_m$ )
4	$A(0) = 1, M(0) = 1 \text{ and } A(1) = 1, M(1) = 0$	Defier
5*	$A(0) = 0, M(0) = 0 \text{ and } A(1) = 0, M(1) = 0$	Never-taker ( $L$ )
6	$A(0) = 0, M(0) = 1 \text{ and } A(1) = 0, M(1) = 1$	Always-taker
7	$A(0) = 0, M(0) = 0 \text{ and } A(1) = 0, M(1) = 1$	Complier
8	$A(0) = 0, M(0) = 1 \text{ and } A(1) = 0, M(1) = 0$	Defier
9	$A(0) = 0, M(0) = 0 \text{ and } A(1) = 1, M(1) = 0$	Complier with respect to moving, never-taker
10	$A(0) = 0, M(0) = 1 \text{ and } A(1) = 1, M(1) = 1$	Complier with respect to moving, always-taker
11*	$A(0) = 0, M(0) = 0 \text{ and } A(1) = 1, M(1) = 1$	Complier with respect to moving, complier ( $R$ )
12	$A(0) = 0, M(0) = 1 \text{ and } A(1) = 1, M(1) = 0$	Complier with respect to moving, defier
13	$A(0) = 1, M(0) = 0 \text{ and } A(1) = 0, M(1) = 0$	Defier with respect to moving, never taker
14	$A(0) = 1, M(0) = 1 \text{ and } A(1) = 0, M(1) = 1$	Defier with respect to moving, always taker
15	$A(0) = 1, M(0) = 0 \text{ and } A(1) = 0, M(1) = 1$	Defier with respect to moving, complier
16	$A(0) = 1, M(0) = 1 \text{ and } A(1) = 0, M(1) = 0$	Defier with respect to moving, defier

Assumption 1*a* eliminates types 13–16 in table B1, assumption 1*b* eliminates type 9, and assumption 1*c* eliminates type 4. This leaves our five types, denoted in the table with an asterisk.

## References

- Abadie, Alberto, and Guido Imbens. 2006. Large sample properties of matching estimators for average treatment effects. *Econometrica* 74, no. 1: 235–67.
- Abdulkadiroglu, Abdulah, Joshua Angrist, Susan Dynarski, Thomas Payne, and Parak Pathak. 2009. Accountability and flexibility in public schools: Evidence from Boston's charters and pilots. Working Paper no. 15549, National Bureau of Economic Research, Cambridge, MA.
- Ahn, Hyuntaik, and James Powell. 1993. Semiparametric estimation of censored selection models with a nonparametric selection mechanism. *Journal of Econometrics* 58:3–29.
- Angrist, Joshua. 1990. Lifetime earnings and the Vietnam era draft lottery: Evidence from Social Security administrative records. *American Economic Review* 80:313–36.
- Angrist, Joshua, Eric Bettinger, Eric Bloom, Elisabeth King, and Michael Kremer. 2002. Vouchers for private schooling in Colombia: Evidence from a randomized natural experiment. *American Economic Review* 92: 1535–58.
- Angrist, Joshua, Guido Imbens, and Donald Rubin. 1996. Identification of causal effects using instrumental variables. *Journal of the American Statistical Association* 91:444–56.
- Barnard, John, Constantine Frangakis, Jennifer Hill, and Donald Rubin. 2003. Principal stratification approach to broken randomized experiments: A case study of school choice vouchers in New York City. *Journal of the American Statistical Association* 98:299–311.
- Bjorklund, Anders, and Robert Moffitt. 1987. The estimation of wage and welfare gains in self-selection models. *Review of Economics and Statistics* 69:42–49.
- Blundell, Richard, Amanda Gosling, Hidehiko Ichimura, and Costas Meghir. 2007. Changes in the distribution of male and female wages accounting for employment composition using bounds. *Econometrica* 75, no. 2:323–63.
- Chernozhukov, Victor, Han Hong, and Eli Tamer. 2007. Estimation and confidence regions for parameter sets in econometric models. *Econometrica* 75, no. 5:1243–84.
- Cullen, Julie, Brian Jacob, and Steven Levitt. 2006. The effect of school choice on student outcomes: Evidence from randomized lotteries. *Econometrica* 74, no. 5:1191–1230.
- Das, Mitali, Whitney Newey, and Francis Vella. 2003. Nonparametric estimation of sample selection models. *Review of Economic Studies* 70:33–58.

- Dobbie, Will, and Roland Fryer. 2009. Are high quality schools enough to close the achievement gap? Evidence from a social experiment in Harlem. Working Paper no. 15473, National Bureau of Economic Research, Cambridge, MA.
- Fisher, Ronald. 1935. *Design of experiments*. New York: Hafner.
- Frangakis, Constantine, and Donald Rubin. 2002. Principal stratification in causal inference. *Biometrics* 58:21–29.
- Frolich, Markus, and Michael Lechner. 2010. Combining matching and nonparametric IV estimation: Theory and an application to the evaluation of active labor market policies. Working paper, <http://ideas.repec.org/p/usg/dp2010/2010-21.html>.
- Hansen, Lars Peter. 1982. Large sample properties of generalized method of moments estimators. *Econometrica* 50, no. 4:1029–53.
- Hastings, Justine, Thomas Kane, and Doug Staiger. 2010. Heterogeneous preferences and the efficacy of public school choice. Working paper, <http://justinehastings.com/research/jh-papers/item/108-hastings-kane-staiger-rev2010>.
- Heckman, James. 1974. Shadow prices, market wages and labor supply. *Econometrica* 42, no. 4:679–94.
- . 1978. Dummy endogenous variables in a simultaneous equation system. *Econometrica* 46:931–60.
- . 1979. Sample selection bias as a specification error. *Econometrica* 47, no. 1:153–61.
- . 1990. Varieties of selection bias. *American Economic Review Papers and Proceedings* 80:313–18.
- Heckman, James, Hidehiko Ichimura, and Petra Todd. 1997. Matching as an econometric evaluation estimator: Evidence from evaluating a job training programme. *Review of Economic Studies* 64, no. 4:605–64.
- Heckman, James, and Richard Robb. 1985. *Alternative methods for evaluating the impact of interventions in longitudinal analysis of labor market data*. Cambridge: Cambridge University Press.
- Heckman, James, Sergio Urzua, and Edward Vytlacil. 2006. Understanding instrumental variables in models with essential heterogeneity. *Review of Economics and Statistics* 88, no. 3:389–432.
- Heckman, James, and Edward Vytlacil. 2005. Structural equations, treatment effects and econometric policy evaluation. *Econometrica* 73, no. 3: 669–738.
- . 2007. Econometric evaluation of social programs. Pts. 1 and 2. In *Handbook of Econometrics*, vol. 6B, ed. James Heckman and Edward Leamer. Amsterdam: Elsevier/North-Holland.
- Horowitz, Joel, and Charles Manski. 2000. Nonparametric analysis of randomized experiments with missing covariate and outcome data. *Journal of the American Statistical Association* 95:77–84.

- Hoxby, Caroline, and Sonali Murarka. 2009. Charter schools in New York City: Who enrolls and how they affect their students' achievement. Working Paper no. 14852, National Bureau of Economic Research, Cambridge, MA.
- Hoxby, Caroline, and Jonah Rockoff. 2005. The impact of charter schools on student achievement. Working paper, <http://www.rand.org/content/dam/rand/www/external/labor/seminars/adp/pdfs/2005hoxby.pdf>.
- Imbens, Guido, and Charles Manski. 2004. Confidence intervals for partially identified parameters. *Econometrica* 72:1845–57.
- Imberman, Scott. 2011. Achievement and behavior in charter schools: Drawing a more complete picture. *Review of Economics and Statistics* 93, no. 2:416–35.
- Lee, David S. 2009. Training, wages, and sample selection: Estimating sharp bounds on treatment effects. *Review of Economic Studies* 76, no. 3:1071–1102.
- Lee, Lung Fei. 1979. Identification and estimation in binary choice models with limited (censored) dependent variables. *Econometrica* 47:977–96.
- Manski, Charles. 1990. Nonparametric bounds on treatment effects. *American Economic Review Papers and Proceedings* 80:319–23.
- Moffitt, Robert. 2008. Estimating marginal treatment effects in heterogeneous populations. *Annals of Economics and Statistics/Annales d'Économie et de Statistique*, nos. 91/92:239–61.
- Neyman, Jerzy. 1923/1990. On the application of probability theory to agricultural experiments: Essay on principles, section 9 [trans. from Polish]. *Statistical Science* 5:465–80.
- Quandt, Richard. 1972. A new approach to estimating switching regression models. *Journal of the American Statistical Association* 67:306–10.
- Rosenbaum, Paul, and Don Rubin. 1983. The central role of the propensity score in observational studies for causal effects. *Biometrika* 70:41–55.
- Rouse, Cecilia. 1998. Private school vouchers and student achievement: An evaluation of the Milwaukee Parental Choice Program. *Quarterly Journal of Economics* 113:553–602.
- Rubin, Don. 1974. Estimating causal effects of treatment in randomized and nonrandomized studies. *Journal of Educational Psychology* 66:688–701.
- . 1976. Inference and missing data. *Biometrika* 63:581–92.
- . 1978. Bayesian inference for causal effects. *Annals of Statistics* 6: 34–58.
- Zhang, Junni, and Don Rubin. 2003. Estimation of causal effects via principal stratification when some outcomes are truncated by “death.” *Journal of Educational and Behavioral Statistics* 28:353–68.
- Zhang, Junni, Don Rubin, and Fabrizia Mealli. 2009. Likelihood-based analysis of causal effects of job-training programs using principal stratification. *Journal of the American Statistical Association* 104:166–76.