Bounding the Treatment Effects of Education Programs That Have Lotteried Admission and Selective Attrition*

John Engberg, Dennis Epple, Jason Imbrogno, Holger Sieg and Ron Zimmer

April 11, 2011

*John Engberg, RAND Corporation, engberg@rand.org. Dennis Epple, Carnegie Mellon, epple@cmu.edu. Jason Imbrogno, Carnegie Mellon, jpi@cmu.edu. Holger Sieg, University of Pennsylvania, holgeres@econ.upenn.edu. Ron Zimmer, Vanderbilt, ronald.w.zimmer@Vanderbilt.Edu. We would like to thank Josh Angrist, Eric Bettinger, Judy Cullen, Stefan Holderlein, Guido Imbens, Blaise Melly, Robert Moffitt, Parag Pathak, Chris Taber, Ken Wolpin, Jeff Wooldridge, Tiemen Woutersen, and seminar participants at Brown University, Carnegie Mellon University, CESifo in Munich, Columbia University, the University of Georgia, Johns Hopkins University, the University of Maryland, Michigan State University, the University of Texas at Austin, the University of Wisconsin-Madison, the 5th Conference of German Economists Abroad at the University of Bonn, and the 2010 ASSA meeting for comments. We would also like to thank the "mid-sized urban school district" for sharing their data. Financial support for this research is provided by the Institute

Abstract

The purpose of this paper is to estimate sharp bounds on treatment effects of education programs that ration excess demand by admission lotteries when selective attrition cannot be ignored. Differential attrition arises in these models because students that lose the lottery are more likely to pursue educational options outside the school district. When students leave the district, important outcome variables are often not observed. Selective attrition implies that treatment effects are not point identified. We provide a new estimator that exploits known quantiles of the outcome distribution to construct informative bounds on treatment effects. We apply our methods to study the effectiveness of magnet programs in a mid-sized urban school district. Our findings show that magnet programs help the district to attract and retain students. The bound estimates demonstrate that magnet programs offered by the district improve behavioral outcomes such as offenses, timeliness, and attendance.

Keywords: Causal Effects, Treatment Effects, Differential Attrition, Noncompliance, Program Evaluation, Randomized Experiments, Instrumental Variables, Magnet Programs, Urban School District, School Choice.

JEL classification: C21, I21, H75

---

of Education Sciences (IES R305A070117 and R305D090016).

1 Introduction

The purpose of this paper is to estimate sharp bounds on treatment effects of education programs that ration excess demand by admission lotteries when selective attrition cannot be ignored. Many school districts use lotteries to determine access to over-subscribed 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. If educational outcomes are not observed for students that leave the school system and attrition rates differ by lottery status, the randomization inherent in the lottery assignment is not necessarily sufficient to identify meaningful treatment effects. Selective attrition may also arise when lottery winners that initially participate in the program drop out because they experience unfavorable outcomes.

The starting point of our analysis is the insight that lotteries can be viewed as experimental designs with multiple sources of non-compliance that arise from parental or student decisions. Since our application focuses on magnet programs, we develop our methods in this context.[1] We focus on two of the most important outside options: parents can send their children to a non-magnet 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 non-compliance with the intended treatment using five latent household types. It is useful to distinguish among these latent types since not all types of non-compliance lead to selective attrition problems. We face different types of missing data problems for different

[1]The methods derived in this paper apply quite broadly to many different educational programs such as charter schools and open enrollment policies.

non-compliers.

The first type is a "complying stayer" that chooses the magnet program if it wins the lottery. The second type is a "non-complying stayer" that does not choose the magnet program even if it wins. Both of these types stay in the district regardless of lottery outcome.[2] The third and fourth types leave the district if they lose the lottery. The third type is a "leaver" and will not enroll its child in the district independently of the outcome of the lottery. The fourth type complies with the lottery and participates in the magnet program if it wins the lottery and leaves if it loses. We denote these households as "at risk," since they are at risk of leaving the district. Given that many urban school districts are experiencing declining enrollment, which affects funding and district programs, this type is important from a policy perspective. Finally, there is a fifth type, the "always takers," that enrolls in the magnet option regardless of the outcome of the lottery.

The household types are latent, i.e. unobserved by both the researcher and the school district administrators.[3] Differential attrition arises in this model due to the presence of "at risk" households for whom we do not observed educational outcomes when they leave. We show how to identify and estimate the proportions of these five latent types. We also characterize differences in observed characteristics among these types. If the households that cause the differential attrition problem differ in observed characteristics from the other latent types, one may also expect that they differ in unobserved characteristics. Our approach thus allows us to characterize the

[2]The district offers a standard education program to all households that do not win the lottery.

[3]Comparing our approach to the one developed in Angrist, Imbens, and Rubin (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" 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. These two assumptions give rise to the differential attrition problem.

extent of the differential attrition problem.

We then discuss how to estimate sharp bounds on the treatment effect of educational programs. The object of interest is the (local) average treatment effect for complying stayers. It is well-known that the standard IV estimator is only consistent if selective attrition can be ignored.[4]

If point identification is not feasible, researchers have typically relied on "worst-case" scenarios to construct bounds for treatment effects. Horowitz and Manski (2000) provide a framework that exploits the assumption that the support of the outcome variable is bounded. Lee (2009) has recently proposed the use of sample trimming rules to construct more informative bounds. The basic idea of his estimator is to assume that the marginal group that only participates because of the treatment is either at the top of the bottom of the observed distribution. Our approach is in the spirit of Lee's, but uses known quantiles of the outcome distribution (test scores at the state level) to create "worst-case" scenarios. Our approach has the advantage that it does not rely on a trimming rule which is helpful when samples are small and power is an issue. Moreover, our estimator allows us to impose all orthogonality conditions that arise from our model simultaneously which can result in significant efficiency gains. This is exhibited by our empirical findings that show that our bound estimates are typically tighter than the ones obtained from the Lee estimator.[5]

Our approach also explicitly deals with heterogeneity in treatment across different schools (or job training centers, as in Lee.) Since estimation is not feasible for

[4]If there are two different types of compliers, the IV estimator does not identity a local average treatment effect. A related paper is Heckman, Urzua, and Vytlacil (2006) who also consider multiple unordered treatments with an instrument shifting agents into one of the treatments.

[5]There are two related papers that use bounding methods. Dinardo, McCrary, and Sanbonmatsu (2006) develops a bounding method that requires an instrument for attrition. Blundell, Gossling, Ichimura and Meghir (2007) develops bounds for the quantiles of the treatment distribution, rather than using an extreme quantile of the outcome distribution to bound the average treatment effect.

each school, researchers often pool data across schools. This creates an aggregation problem in estimation. Our estimators deal with the aggregation problem that is encountered when researchers have to pool among lotteries to deal with small sample problems. We show that flexible weighting schemes can be employed to estimate meaningful weighted averages of the underlying mean treatment effects.

We apply the techniques developed in this paper to study the effectiveness of magnet programs in a mid-sized urban school district. A second contribution of this paper is that we provide new research to understand the causal effects of magnet programs. While debates surrounding the effectiveness of other school choice options such as charter schools and educational vouchers have attracted much attention from researchers and policymakers, magnet programs have gotten less attention despite the fact that they are much more prevalent than charter schools or educational voucher programs.

Our findings show that magnet programs help the district to attract and retain students. Approximately 25 percent of applicants to magnet programs that serve K-5 students are "at risk." Thus selective attrition poses an important problem for the school district in our application. Households that selectively attrit come from neighborhoods that have higher incomes and are more educated than households that stay in the district regardless of the outcome of the lottery. These "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. We also find that the market for elementary school education is more competitive than the market for middle and high school education. The fraction of households at risk declines with the age of the students.

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 only conducted in grades 5, 8, and 11 during most of our sample period. For a variety of behavioral outcomes, we do not face these data limitations. 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 paper is related to a growing literature that evaluates educational programs using lottery based estimators.[6] Lotteries were used by Rouse (1998) to study the impact of the Milwaukee voucher program. Angrist, Bettinger, Bloom, King, and Kremer (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 (2004) use lotteries to study Chicago charter schools. Cullen, Jacob, and Levitt (2006) have analyzed open enrollment programs in the Chicago Public Schools. Ballou, Goldring, and Liu (2006) examine a magnet program. Hastings, Kane, and Staiger (2008) estimate a model of school choice based on stated preferences for schools in Charlotte. 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, Angrist, Dynarski, Payne, and Pathak (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 achievement gains. All of these papers focus on applications in which selective attrition is not present and thus do not explicitly deal with the key selective attrition problem discussed in this paper.[7]

[6] Angrist (1990) introduced the use of lotteries to study the impact of military service on earnings.

[7] 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.

The rest of the paper is organized as follows. Section 2 develops our new methods for estimation of treatment effects when program participation is partially determined by lotteries and selective attrition cannot be ignored. We discuss identification and estimation. Section 3 provides some institutional background for our application and discusses our main data sources. Section 4 reports the empirical findings of our paper. Finally, we offer some conclusions and discuss the policy implications of our work in Section 5.

2 Identification and Estimation

2.1 The Research Design

We consider a design that arises when randomization determines eligibility to participate in an educational program. Consider the problem of a parent that has to decide whether or not to enroll a student in a magnet program offered by a school district.[8] We only consider households that participate in a lottery that determines access to an oversubscribed (magnet) program. Let W denote a discrete random variable which is equal to 1 if the student wins the lottery and 0 if it 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, non-magnet 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 1 if a student attends the (magnet) program and 0 otherwise. Finally, let A denote a random variable that is 1 if a student attends a school in the

[8] We use the terms "parent" or "households" to describe the decision maker and "student" to describe the person that participates in the program.

district and 0 otherwise.

To model compliance with the intended treatment, we use five latent types to classify households into compliers and non-compliers. We make the following assumption.

Assumption 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 non-magnet 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.*[9]

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.*

[9]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.

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, one may also expect that they 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 variable X that measures an observed household characteristic such as income or socio-economic status. Appealing to our decomposition, let μ_r, μ_{s_m}, μ_{s_n}, μ_l and μ_{a_t} denote the means of random variable X conditional on belonging to group r, s_m, s_n, l, and a_t, respectively. The goal of the first part of the analysis is then to identify and estimate the following eleven parameters $(w, r, s_n, s_m, l, a, \mu_r, \mu_{s_n}, \mu_{s_m}, \mu_l, \mu_{a_t})$.[10]

The next objective is to study the effects of the program on student outcomes. Let T be an outcome measure of interest, for example, the score on a standardized achievement test. Following Fisher (1935), we adopt standard notation in the program evaluation literature and consider a model with three potential outcomes:

$$T \;=\; A\, M\, T_1 \;+\; A\,(1-M)\, T_0 \;+\; (1-A)\, T_2 \tag{1}$$

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

[10] It is straightforward to allow X to be a vector.

of the district.[11] We will later assume that T is not observed for students that do not attend a public school within the district, i.e. iT_2 is not observed. This assumption is plausible since researchers typically only have access to data from 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 Δ is unobserved for all students. Conceptually, we can define five different average treatment effects, one for each latent group.[12]

$$ATE_{Type} \;=\; E[T_1 - T_0 | Type = 1] \qquad Type \in \{S_n, S_m, R, L, A_t\} \tag{2}$$

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.

2.2 Identification of the Fraction of Latent Types

First we need to establish the information set of the researcher.

Assumption 2 *The researcher observes probabilities and conditional means for the feasible outcomes shown in Table 1.*

[11] 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 Vytlacil (2007) for an overview of the program evaluation literature.

[12] 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 Vytlacil (2005) and Moffitt (2008).

Table 1: Observed Outcomes

	W	M	A	$Pr\{W, M, A\}$	$E[X\|W, M, A] = E[X\|M, A]$
I	1	1	1	$w\ (r + s_m + a_t)$	$\frac{r\mu_r + s_m\mu_{s_m} + a_t\mu_{a_t}}{r + s_m + a_t}$
II	1	1	0	not possible	
III	1	0	1	$w\ s_n$	μ_{s_n}
IV	1	0	0	$w\ l$	μ_l
V	0	1	1	$(1 - w)a_t$	μ_{a_t}
VI	0	1	0	not possible	
VII	0	0	1	$(1 - w)\ (s_n + s_m)$	$\frac{s_n\mu_{s_n} + s_m\mu_{s_m}}{s_n + s_m}$
VIII	0	0	0	$(1 - w)\ (r + l)$	$\frac{r\mu_r + l\mu_l}{r + l}$

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) *are identified by the six non-degenerate probabilities in Table 1.*

Proof: Parameter w is the fraction that wins the lottery:

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

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

$$s_n = Pr(W = 1, M = 0, A = 1)/w \tag{4}$$

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

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

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

$$a_t \;\; = \;\; Pr(W=0, M=1, A=1)/(1-w) \tag{6}$$

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

$$s_m \;\; = \;\; Pr(W=0, M=0, A=1)/(1-w) \; - \; s_n \tag{7}$$

Given a_t, l, s_n, and s_m, r is identified of the identity:

$$r \;\; = \;\; 1 - l - s_m - s_n - a_t \tag{8}$$

Q.E.D.

Note that there is no over-identification at this stage since the six probabilities in Table 1 add up to one, and the last three non-degenerate 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 observed in Table 1.*

Proof: μ_l is identified from (1,0,0):

$$\mu_l \;\; = \;\; E(X|W=1, M=0, A=0) \tag{9}$$

Similarly μ_{s_n} is identified from (1,0,1):

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

and μ_{a_t} is identified from (0,1,1):

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

Given μ_{s_n}, μ_{s_m} is identified from (0,0,1):

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

Given μ_{s_m} and μ_{a_t}, μ_r is identified from (1,1,1):

$$\mu_r \;=\; [(r + s_m + a_t)E(X|W=1, M=1, A=1) \;-\; s_m\mu_{s_m} \;-\; a_t\mu_{a_t}]/r \tag{13}$$

Q.E.D.

There is one over-identifying condition at this stage. This restriction arises due to the condition that W is orthogonal to X.[13] 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. We show this in the next section.

2.3 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 only observes outcomes, T, for students that remain in the school district, i.e. we do not observe outcomes for "leavers" and "at risk" households that lose the lottery.

[13] The lotteries are assumed to be fair and blind in the sense that the district does not pre-select winners and losers based on beliefs about attendance or any socio-economic or student characteristic found in X.

It is useful to assume initially that we observe the latent household type. Table 2 provides a summary of the relevant conditional expectations.[14] Conditioning on lottery outcomes, there are ten 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.

Table 2: Mean Outcomes Conditional on Type

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

Note that T_2 is never observed.

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 that 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 can neither identify $E[T_1|L = 1]$ nor $E[T_0|L = 1]$. 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}. 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

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

not straightforward. One key result of this paper is that the local average treatment effect for compliers is not point identified if there is selective attrition.

Proposition 3

If there is selective attrition ($r \neq 0$) and if households that are at risk have different expected outcomes than compliers in the treated case ($E[T_1|S_m = 1] \neq E[T_1|R = 1]$), then the local average treatment effect for compliers, ATE_{S_m}, is not identified.

Proof:

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

$$E[T|W = 1, M = 1, A = 1] \;=\; \frac{s_m E[T_1|S_m = 1] + rE[T_1|R = 1] + a_t E[T_1|A_t = 1]}{s_m + r + a_t} \tag{14}$$

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

$$E[T|W = 0, M = 1, A = 1] \;=\; E[T_1|A_t = 1] \tag{15}$$

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

$$E[T|W = 0, M = 0, A = 1] \;=\; \frac{s_m E[T_0|S_m = 1] + s_n E[T_0|S_n = 1]}{s_m + s_n} \tag{16}$$

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

$$E[T|W = 1, M = 0, A = 1] \;=\; E[T_0|S_n = 1] \tag{17}$$

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 $E[T_1|R = 1]$. Q.E.D.

Proposition 3 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.[15] The lack of point identification arises from the "at risk" households which cause the selective attrition problem. Selective attrition is only a problem if "at risk" households have different mean outcomes than compliers.[16]

Since point identification is no longer feasible when selective attrition is not negligible, attention focuses on set identification and the construction of bounds.[17]

Proposition 4

i) Suppose we have an upper bound, denoted by T_1^u, *for* $E[T_1|R = 1]$ *i.e.* T_1^u *satisfies* $E[T_1|R = 1] \leq T_1^u$. *We can then construct a lower bound for the* $E[T_1|S_m = 1]$ *and* ATE_{S_m}.

ii) Suppose we have a lower bound, denoted by T_1^l, *for* $E[T_1|R = 1]$, *i.e.* T_1^l *satisfies* $E[T_1|R = 1] \geq T_1^l$, *we can then construct an upper bound for the* $E[T_1|S_m = 1]$ *and* ATE_{S_m}.

Proof:

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

$$E[T_1|S_m = 1]$$

[15]Recall that if $r = l = 0$ our research design simplifies to the one considered in Angrist, Imbens and Rubin (1996).

[16]We can generalize Proposition 3 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).

[17]Point identification cannot be achieved in many econometric applications. In that case, attention naturally shifts to characterizing informative bounds on the parameters of interest. See, for example Manski (1997), Horowitz and Manski (2000) Imbens and Manski (2004), Chernozhukov, Imbens, and Newey (2006), and Lee (2009).

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

where the last inequality follows from $E[T_1|R=1] \leq T_1^u$. Since all terms in the last line of equation (18) are identified, we conclude that we can construct a lower bound. Replacing T_1^u with T_1^l and reversing the inequality yields the upper bound. Q.E.D.

There are many ways of constructing both lower bounds or upper bounds depending on the outcome variable and the application. For example, 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 paper is to bound outcomes using known percentiles of the outcome distribution. These type 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 can apply the trimming approach 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 based both on the proportions of missing data in the control and treatment groups and the distribution of the outcome variable being bounded.

We have thus 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 we cannot reject the null hypothesis that $r = 0$, treatment effects are point identified and can be estimated using standard linear 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 that 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 paper is more appropriate than IV estimation.

2.4 A GMM Estimator

Suppose 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. Nevertheless, 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 based on moment conditions derived from the choice probabilities in Table 1. Define:

$$f_1(A_i, M_i, W_i) = \begin{cases} \frac{1}{N}\sum_{i=1}^{N}\left[W_i M_i A_i - w(r + s_m + a_t)\right] \\ \frac{1}{N}\sum_{i=1}^{N}\left[W_i(1-M_i)A_i - w\ s_n\right] \\ \frac{1}{N}\sum_{i=1}^{N}\left[W_i(1-M_i)(1-A_i) - w\ l\right] \\ \frac{1}{N}\sum_{i=1}^{N}\left[(1-W_i)M_i A_i - (1-w)\ a_t\right] \\ \frac{1}{N}\sum_{i=1}^{N}\left[(1-W_i)(1-M_i)A_i - (1-w)(s_n + s_m)\right] \end{cases}$$

and note that $E[f_1(A_i, M_i, W_i)] = 0$. Similarly we can estimate the mean characteristics of each type. Define:

$$f_2(A_i, M_i, W_i, X_i) \quad = \quad \begin{cases} \frac{1}{N}\sum_{i=1}^{N}\Big[W_iM_iA_iX_i - w[r\mu_r + s_m\mu_{s_m} + a_t\mu_{a_t}]\Big] \\ \frac{1}{N}\sum_{i=1}^{N}\Big[W_i(1-M_i)A_iX_i - w\ s_n\ \mu_{s_n}\Big] \\ \frac{1}{N}\sum_{i=1}^{N}\Big[W_i(1-M_i)(1-A_i)X_i - w\ l\ \mu_l\Big] \\ \frac{1}{N}\sum_{i=1}^{N}\Big[(1-W_i)M_iA_iX_i - (1-w)a_t\ \mu_{a_t}\Big] \\ \frac{1}{N}\sum_{i=1}^{N}\Big[(1-W_i)(1-M_i)A_iX_i - (1-w)[s_n\mu_{s_n} + s_m\mu_{s_m}]\Big] \\ \frac{1}{N}\sum_{i=1}^{N}\Big[(1-W_i)(1-M_i)(1-A_i)X_i - (1-w)[r\mu_r + l\mu_l]\Big] \end{cases}$$

and note that $E[f_2(A_i, M_i, W_i)] = 0$. 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 $E[T_1|R = 1]$ to the 5th percentile of the observed outcome distribution, denoted by T_1^l. Define:

$$f_3(A_i, M_i, W_i, T_i) = \begin{cases} \frac{1}{N}\sum_{i=1}^{N}\Big[T_iW_iM_iA_i - w(s_mE[T_1^u|S_m = 1] + rT_1^l + a_tE[T_1|A_t = 1])\Big] \\ \frac{1}{N}\sum_{i=1}^{N}\Big[T_i(1-W_i)M_iA_i - (1-w)a_tE[T_1|A_t = 1]\Big] \\ \frac{1}{N}\sum_{i=1}^{N}\Big[T_i(1-W_i)(1-M_i)A_i - (1-w)(s_mE[T_0|S_m = 1] + s_nE[T_0|S_n = 1])\Big] \\ \frac{1}{N}\sum_{i=1}^{N}\Big[T_iW_i(1-M_i)A_i - ws_nE[T_0|S_n = 1]\Big] \end{cases}$$

and we have $E[f_3(A_i, M_i, W_i)] = 0$. 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 non-missing data for all other outcomes. Combining all orthogonality conditions, we can estimate the parameters of the model using a GMM estimator (Hansen, 1982). Note that the estimator above easily generalizes to the case in which X is a vector of random variables. We simply stack all orthogonality conditions to obtain a simultaneous estimator. 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. Many of the parameters of the model – especially all parameters that characterize the fraction of latent types – can be estimated using linear estimators.[18] We find in the application that imposing the additional orthogonality conditions that model the mean characteristics of the types ($f_2(A_i, M_i, W_i, X_i)$ above) yields significant efficiency gains.

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 paper 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 to the objectives of the policy or decision maker.

3 Data

Our application focuses on magnet programs that are operated by a mid-sized urban school district that prefers to not be identified. Magnet schools emerged in the United States in the 1960's. Magnet schools are designed to draw students from across normal attendance zones. In contrast, a feeder school typically only admits students that 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 thus 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 better socio-economic backgrounds in

[18] An appendix is available upon request which shows exactly how to set up the linear estimators.

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

In the case of over-subscription, 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 cut-off number that separates the groups. We do not observe students attending magnet schools that lose the lottery, i.e. 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. When winners decline admission, the students on the wait list become eligible. Again the rank on the wait list is determined by the original lottery. 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 both on observable and unobservable characteristics.

The district granted us access to its longitudinal student database. We use data from the 1999-2000 school year through 2005-2006. In addition to demographic data, the database contains detailed information about educational outcomes. This information is linked to each student by a unique ID number. The demographic characteristics for the students include race, gender, free/reduced lunch eligibility, and addresses.[19] Using the addresses, we can assign census tract level variables to each student. We use two community characteristics that measure the socio-economic composition of the neighborhoods in which students 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 database also contains the outcomes of the magnet lotteries. One of the key features of the database is that it contains unusually good

[19]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.

Table 3: Descriptive Statistics

Variable	*Entire Sample (2054 obs)*	*Elem School (820 obs)*	*Middle School (457 obs)*	*High School (777 obs)*
Gender	0.51 (0.50)	0.51 (0.50)	0.51 (0.50)	0.51 (0.50)
Race	0.75 (0.44)	0.59 (0.49)	0.79 (0.40)	0.88 (0.32)
FRL	0.33 (0.47)	0.33 (0.47)	0.35 (0.48)	0.32 (0.47)
Poverty	0.23 (0.14)	0.22 (0.14)	0.23 (0.14)	0.24 (0.15)
Education	0.29 (0.19)	0.34 (0.22)	0.28 (0.18)	0.25 (0.14)
Offenses	0.99 (2.23)	0.18 (0.99)	1.15 (2.32)	1.67 (2.71)
Suspension Days	1.88 (4.71)	0.29 (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

information about students residing in the district that attend private, charter, and home schools. Unfortunately, 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 sub-samples that we also consider in estimation.[20] We only consider binding lotteries in this 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 non-randomness, 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:

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

where X_i is the observable characteristic of interest, W_i is a dummy equal to 1 if student i wins lottery j, I_{ij} is an indicator variable equal to 1 if student i participated in lottery j, and v_i is the error term.[21] We estimate a separate regression for each observable. The coefficient β_1 determines the fairness of the lottery system. If we

[20] For a small sample of students we imputed absences and tardies. Also note that outcome variables are not observed for students that leave the district. Thus the means of the outcome variables in Table 3 reflect means of stayers.

[21] Alternatively we could use multivariate Behrens-Fisher type test statistics which require less restrictive assumptions. See, for example, Kim (1992).

Table 4: Lottery Balance Result

Variable	*Entire Sample*	*Elem School*	*Middle School*	*High School*
Gender	0.0053	0.0366	-0.0183	-0.0257
	(0.0262)	(0.0384)	(0.0559)	(0.0469)
FRL	0.0056	0.0111	-0.0501	0.0385
	(0.0229)	(0.0322)	(0.0482)	(0.0431)
Poverty	-0.0050	-0.0023	0.0044	-0.0160
	(0.0068)	(0.0092)	(0.0136)	(0.0135)
Education	0.0041	0.0110	-0.0038	-0.0007
	(0.0078)	(0.0127)	(0.0165)	(0.0125)

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.

The first column of Table 4 shows the results when all students in all binding lotteries are included in the regressions. β_1 is not significant for any tested variable at 10 %. The second and third columns consider the three sub-samples of interest. The second column includes all students in elementary school while the third column focuses on middle school students and the fourth on high school students. We find that the estimates of β_1 are not significantly different from 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 sub-samples that were tested.

4 Empirical Results

4.1 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 only imposes the orthogonality conditions that identify the fraction of latent household types. Then we add the orthogonality conditions that capture 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 U.S. Census. We report estimates for three samples which include all students that applied to an oversubscribed magnet program that is associated with an elementary school (ES), middle school (MS), and high school (HS), respectively. We pool across all lotteries in each sample and, therefore, use the weighted estimator discussed in Appendix A. Tables 5 and 6 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 5 clearly allows us to evaluate whether there are efficiency gains that arise when using a GMM estimator.[22] 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 percent larger when one ignores the additional orthogonality conditions. We thus conclude that our approach of jointly estimating the model using GMM is preferable to simpler methods.

Table 5 reveals some interesting new insights into the importance of selective attrition in our application. Recall that the fraction of households at risk is the

[22]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 5: Empirical Results: Selective Attrition

	First Set of Orthogonality Conditions			
	Fraction At Risk	Fraction Stay, Attend	Fraction Stay, Non	Fraction Leave
ES	0.25 (0.04)	0.61 (0.05)	0.06 (0.01)	0.08 (0.01)
MS	0.12 (0.15)	0.60 (0.16)	0.24 (0.04)	0.04 (0.01)
HS	0.15 (0.09)	0.70 (0.09)	0.08 (0.01)	0.06 (0.01)
	First and Second Set of Orthogonality Conditions			
	Fraction At Risk	Fraction Stay, Attend	Fraction Stay, Non	Fraction Leave
ES	0.25 (0.04)	0.61 (0.04)	0.06 (0.01)	0.08 (0.01)
MS	0.12 (0.05)	0.61 (0.06)	0.24 (0.04)	0.04 (0.01)
HS	0.14 (0.06)	0.72 (0.06)	0.08 (0.01)	0.06 (0.01)

Estimated standard errors are reported in parentheses.

key parameter that captures selective attrition. We find that selective attrition is substantial and ranges between 12 and 25 percent 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 to 71 percent, will attend the magnet program if they win they lottery. The fraction of households that will leave the district regardless of the outcome of the lottery ranges between 4 and 8 percent. 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 6. These and the ones reported in the lower part of Table

Table 6: Empirical Results: Characteristics

Gender				
	At Risk	Stay, Attend	Stay, Non	Leave
ES	0.57 (0.09)	0.47 (0.03)	0.55 (0.11)	0.47 (0.08)
MS	0.85 (0.34)	0.43 (0.06)	0.50 (0.08)	0.31 (0.13)
HS	0.55 (0.34)	0.57 (0.05)	0.49 (0.08)	0.41 (0.08)
Race				
	At Risk	Stay, Attend	Stay, Non	Leave
ES	0.50 (0.09)	0.70 (0.04)	0.39 (0.11)	0.18 (0.07)
MS	0.99 (0.41)	0.80 (0.05)	0.80 (0.06)	0.28 (0.14)
HS	0.89 (0.41)	0.93 (0.03)	0.85 (0.07)	0.79 (0.06)
FRL				
	At Risk	Stay, Attend	Stay, Non	Leave
ES	0.12 (0.04)	0.43 (0.03)	0.19 (0.07)	0.07 (0.04)
MS	0.26 (0.15)	0.47 (0.06)	0.26 (0.09)	0.07 (0.06)
HS	0.15 (0.11)	0.39 (0.04)	0.25 (0.06)	0.12 (0.05)
Poverty				
	At Risk	Stay, Attend	Stay, Non	Leave
ES	0.21 (0.03)	0.23 (0.01)	0.20 (0.04)	0.14 (0.01)
MS	0.24 (0.10)	0.24 (0.02)	0.23 (0.02)	0.13 (0.02)
HS	0.28 (0.12)	0.25 (0.01)	0.24 (0.02)	0.19 (0.02)
Education				
	At Risk	Stay, Attend	Stay, Non	Leave
ES	0.40 (0.05)	0.29 (0.02)	0.41 (0.05)	0.53 (0.04)
MS	0.20 (0.11)	0.29 (0.02)	0.30 (0.03)	0.55 (0.08)
HS	0.27 (0.14)	0.25 (0.01)	0.21 (0.02)	0.36 (0.03)

Estimated standard errors are reported in parentheses.

5 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 on free or reduced lunch programs than households that are stayers. Moreover, they come from better educated neighborhoods.[23] 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 back-up option in case their students should unexpectedly not be admitted to an independent, charter, or parochial school.[24]

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 "at risk" students are only as good as the "stayers," while the lower bound estimation assumes that the at risk students are in the 95th percentile of the outcome distribution.

Table 6 also permits interesting comparisons across grade levels. Elementary and middle school lotteries are somewhat more competitive than high school lotteries. The former have average win rates of 52 percent and 53 percent respectively while the latter have an average win rate of 77 percent. 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

[23]Note that the differences in household characteristics are statistically significant from zero at all conventional levels.

[24]It could also be that these households left the district because of job transfers or other issues unrelated to schools.

find that the fraction of at risk families and the fraction of leavers is 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.

4.2 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 socio-economic 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 only set-identified when selective attrition matters, we implement our bounds estimators. We implement our bounds estimators by adding the orthogonality conditions for these variables to the conditions, discussed in Section 4.1, for estimating the proportions of latent types and the demographic characteristics of latent types. 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 due to missing data. This is largely the case because standardized achievement tests were only conducted 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 test scores. For the high school sample, the reduction is of similar magnitude.[25] Including households that participate in the lotteries but subsequently leave the district gives us with 213 middle school students and 203 high

[25]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.

school students. Table 7 summarizes our main findings using standardized test scores in reading and mathematics as outcome variables.

Table 7: Empirical Results: Achievement

	Reading			Mathematics		
	Upper Bound ATE_{S_m}	Lower Bound ATE_{S_m}	IV ATE_{S_m}	Upper Bound ATE_{S_m}	Lower Bound ATE_{S_m}	IV ATE_{S_m}
MS	66.25	3.68	139.71	180.89	91.08	138.56
	(118.30)	(172.69)	(77.33)	(124.89)	(183.69)	(63.63)
HS	77.05	-25.09	81.97	87.09	-24.22	94.30
	(64.79)	(136.24)	(47.17)	(57.62)	(148.00)	(40.70)

Estimated standard errors are reported in parentheses.

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 people 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 reveal that this inference is not correct.

We next turn our attention to behavioral outcomes measured one year after the lotteries were conducted.[26] 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,

[26]Previously Cullen et al. (2006) and Imberman (2010) have studied behavioral outcomes when examining school choice programs.

Table 8: Empirical Results: Behavioral Outcomes

	Offenses			Suspensions		
	Upper Bound ATE_{S_m}	Lower Bound ATE_{S_m}	IV ATE_{S_m}	Upper Bound ATE_{S_m}	Lower Bound ATE_{S_m}	IV ATE_{S_m}
ES	-0.28 (0.09)	-0.26 (0.09)	-0.26 (0.09)	-0.49 (0.15)	-0.45 (0.15)	-0.47 (0.14)
MS	-0.62 (0.36)	-0.48 (0.36)	-0.66 (0.35)	-0.22 (1.17)	0.00 (1.18)	-0.56 (0.77)
HS	-0.03 (0.34)	0.28 (0.39)	0.20 (0.31)	-0.47 (0.87)	0.14 (0.93)	0.03 (0.75)
	Absences			Tardies		
	Upper Bound ATE_{S_m}	Lower Bound ATE_{S_m}	IV ATE_{S_m}	Upper Bound ATE_{S_m}	Lower Bound ATE_{S_m}	IV ATE_{S_m}
ES	-2.26 (0.90)	0.98 (1.24)	-1.70 (0.77)	-0.95 (0.73)	0.52 (0.87)	-0.98 (0.59)
MS	1.98 (1.60)	4.16 (2.02)	1.82 (1.36)	3.04 (1.82)	4.97 (2.07)	2.32 (2.07)
HS	-8.64 (3.32)	-5.35 (3.60)	-7.77 (2.55)	-7.90 (2.78)	-6.61 (2.87)	-9.41 (2.45)

Estimated standard errors are reported in parentheses.

absences, and tardies.

Table 8 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 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 signicantly 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 the behavioral outcomes is zero. Thus our lower bound estimates for all behavioral outcomes is the most pessimistic possible, since it attributes flawless behavior to all who leave the district.

4.3 Comparison with the Lee Estimator

The main alternative to our estimator is the one proposed by Lee (2009) that relies on trimming to construct an estimator for the lower and upper bounds of the treatment effect. It is, therefore, useful to compare both approaches using the data from our application. Table 9 compares our estimates with those obtained from Lees trimming method.[27] As we detail in the appendix, weighting is appropriate when estimating bounds using data from multiple lotteries. In implementing Lees estimator, we do not weight lotteries by number of applicants.[28] Hence, the comparison in Table 9 reflects both a difference in the approach to bounding as well as a difference in weighting, potentially confounding the two effects. For the outcomes considered in Table 9, we have confirmed that the results from our weighted estimator are similar to those when we do not weight by lotteries. This is not always the case, however. For example, for MS reading, weighting by lotteries proves to be quite important.[29] 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 9 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. Table 9 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. The TE CI is the treatment effect confidence interval.

We find that the trimming rates are much greater in our application than in Lees application, where $\hat{p} = 0.068$. This is due to the fact that our proportion

[27]The results are similar for other outcomes analyzed in this paper. 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.

[28]Lees estimator has not yet been extended to estimate bounds when combining data from multiple lotteries, though it is surely possible to do so.

[29]Details are available on request.

Table 9: Comparison with Lee Estimator

	Our Estimator	Lee's Estimator
ES Offenses	UB : -0.28 (0.09) LB : -0.26 (0.09) Point Estimate Range : 0.02 Simple TE CI : [-0.46 , -0.08]	UB : -0.33 (0.08) [347] LB : -0.27 (0.08) [357] Point Estimate Range : 0.06 Simple TE CI : [-0.49 , -0.11] $\hat{p} = 0.337$
ES Tardies	UB : -0.95 (0.73) LB : 0.52 (0.87) Point Estimate Range : 1.47 Simple TE CI : [-2.38 , 2.23]	UB : -3.58 (0.58) [217] LB : -0.68 (0.74) [306] Point Estimate Range : 2.90 Simple TE CI : [-4.72 , 0.77] $\hat{p} = 0.362$
HS Math	UB : 87.09 (57.62) LB : -24.22 (148.00) Point Estimate Range : 111.31 Simple TE CI : [-314.30 , 200.03]	UB : 243.69 (301.97) [33] LB : -150.83 (252.33) [33] Point Estimate Range : 394.52 Simple TE CI : [-645.40 , 835.55] $\hat{p} = 0.660$
MS Math	UB : 180.89 (124.89) LB : 91.08 (183.69) Point Estimate Range : 89.81 Simple TE CI : [-268.95 , 425.67]	UB : 382.86 (286.89) [45] LB : 65.11 (243.81) [48] Point Estimate Range : 317.75 Simple TE CI : [-412.76 , 945.16] $\hat{p} = 0.426$

of non-missing 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 3000 observations in the treatment group before and after trimming. These sample are much larger than the ones in our application. Trimming can, therefore, lead to small sample estimation problems in some applications.

5 Conclusions

We have considered a research design that arises when randomization is used to determine access to oversubscribed programs offered by public school systems. We have developed a new empirical method which deals with selective attrition. Our approach classifies potential participants as stayers, always-takers, leavers, and those that are at risk. We show that the last type of households causes the selective attrition problem. These "at risk" households are also most interesting from a policy perspective since the decision to remain in public schooling crucially depends on the outcome of the lottery. If selective attrition matters, point identification of local average treatment effects for compliers cannot be established. Instead we show how to construct and estimate informative bounds.

We have applied our new methods to study the effectiveness of magnet programs. 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 believe that the techniques discussed in this paper can be extended and applied to variety of different problems. Chan and Hamilton (2006), for example, consider clinical AIDS trials and show that attrition is prevalent. Dinardo et al. (2006) show that attrition is also a problem in the Moving To Opportunity randomized experiment. The techniques developed in this paper can be applied to study these types of questions as well.

References

Abadie, A. and Imbens, G. (2006). Large Sample Properties of Matching Estimators for Average Treatment Effects. *Econometrica*, *74 (1)*, 235–67.

Abdulkadiroglu, A., Angrist, Dynarski, S., Payne, T., and Pathak, P. (2009). Accountability and Flexibility in Public Schools: Evidence from Boston's Charters and Pilots. NBER Working Paper 15549.

Angrist, J. (1990). Lifetime Earnings and the Vietnam Era Draft Lottery: Evidence from Social Security Administrative Records. *American Economic Review*, *80*, 313–36.

Angrist, J., Bettinger, E., Bloom, E., King, E., and Kremer, M. (2002). Vouchers for Private Schooling in Colombia: Evidence from a Randomized Natural Experiment. *American Economic Review*, *92*, 1535–58.

Angrist, J., Bettinger, E., and Kremer, M. (2005). Long-Term Educational Consequences of Secondary School Vouchers: Evidence from Administrative Records in Colombia. Working Paper.

Angrist, J., Imbens, G., and Rubin, D. (1996). Identification of Causal Effects Using Instrumental Variables. *Journal of the American Statistical Association*, *91*, 444–56.

Ballou, D., Goldring, E., and Liu, K. (2006). Magnet Schools and Student Achievement. Working Paper.

Bjorklund, A. and Moffitt, R. (1987). The Estimation of Wage and Welfare Gains in Self-Selection Models. *Review of Economics and Statistics*, *69*, 42–49.

Blundell, R. Gossling, A., Ichimura, H. and Meghir, C. (2007). Changes in the Distribution of Male and Female Wages Accounting for Employment Composition Using Bounds. *Econometrica*, *75 (2)*, 323-63.

Chan, T. and Hamilton, B. (2006). Learning, Private Information, and the Economic Evaluation of Randomized Experiments. *Journal of Political Economy*, *114 (6)*, 997-1040.

Chernozhukov, V., Imbens, G., and Newey, W. (2006). Nonparametric Identification and Estimation of Non-separable Models. *Journal of Econometrics*, *forthcoming*.

Cullen, J., Jacob, B., and Levitt, S. (2006). The Effect of School Choice on Student Outcomes: Evidence From Randomized Lotteries. *Econometrica*, *74(5)*, 1191–230.

, Dinardo, J., McCrary, J., and Sanbonmatsu, L. (2006). Constructive Proposals for Dealing with Attrition: An Empirical Example.. NBER Working Paper.

Dobbie, W. and Fryer, R. (2009). Are High Quality Schools Enough to Close the Achievement Gap? Evidence From a Social Experiment in Harlem.. NBER Working Paper 15473.

Fisher, R. (1935). *Design of Experiments*. Hafner. New York.

Hansen, L. P. (1982). Large Sample Properties of Generalized Method of Moments Estimators. *Econometrica*, *50 (4)*, 1029–1053.

Hastings, J., Kane, T., and Staiger, D. (2008). Heterogeneous Preferences and the Efficacy of Public School Choice. Working Paper.

Heckman, J. (1978). Dummy Endogenous Variables in a Simultaneous Equation System. *Econometrica*, *46*, 931–960.

Heckman, J. (1979). Sample Selection Bias as a Specification Error. *Econometrica*, *47 (1)*, 153–161.

Heckman, J., Ichimura, H., and Todd, P. (1997). Matching As An Econometric Evaluation Estimator: Evidence from Evaluating a Job Training Program. *Review of Economic Studies*, *64 (4)*.

Heckman, J. and Robb, R. (1985). Alternative Methods for Evaluating the Impact of Interventions. In *Longitudinal Analysis of Labor Market Data*. Cambridge University Press.

Heckman, J., Urzua, S., and Vytlacil, E. (2006). Understanding Instrumental Variables in Models with Essential Heterogeneity. *Review of Economics and Statistics*, *88 (3)*, 389–432.

Heckman, J. and Vytlacil, E. (2005). Structural Equations, Treatment Effects and Econometric Policy Evaluation. *Econometrica*, *73*.

Heckman, J. and Vytlacil, E. (2007). Econometric Evaluation of Social Programs. In *Handbook of Econometrics 6b*. Elsevier North Holland.

Horowitz, J. and Manski, C. (2004). Nonparametric Analysis of Randomized Experiments With Missing Covariate and Outcome Data. *Journal of the American Statistical Association*, *95*, 77-84.

Hoxby, C. and Murarka, S. (2009). Charter Schools in New York City: Who Enrolls and How they Affect Their Students' Achievement. NBER Working Paper 14852.

Hoxby, C. and Rockoff, J. (2004). The Impact of Charter Schools on Student Achievement. Working Paper.

Imbens, G. and Manski, C. (2004). Confidence Intervals for Partially Identified Parameters. *Econometrica*, *72*, 1845–1857.

Imberman, S. (2010). Achievement and Behavior in Charter Schools: Drawing a More Complete Picture. *Review of Economics and Statistics*, *forthcoming*.

Kim, S. (1992). A Practical Solution to the Multivariate Behrens-Fisher Problem. *Biometrica*, *79*, 171–6.

Lee, L. (1979). Identification and Estimation in Binary Choice Models with Limited (Censored) Dependent Variables. *Econometrica*, *47*, 977–996.

Lee, D. (2009). Training, Wages, and Sample Selection: Estimating Sharp Bounds on Treatment Effects. *Review of Economic Studies*, *76(3)*, 1071–1102.

Manski, C. (1997). Monotone Treatment Response. *Econometrica*, *65*, 1311–1334.

Moffitt, R. (2008). Estimating Marginal Treatment Effects in Heterogeneous Populations. Working Paper.

Quandt, R. (1972). A New Approach to Estimating Switching Regression Models. *Journal of the American Statistical Association*, *67*, 306–310.

Rosenbaum, P. and Rubin, D. (1983). The central role of the propensity score in observational studies for causal effects. *Biometrica*, *70*, 41–55.

Rouse, C. (1998). Private School Vouchers and Student Achievement: An Evaluation of the Milwaukee Parental Choice Program. *Quarterly Journal of Economics*, *113*, 553–602.

Rubin, D. (1974). Estimating Causal Effects of Treatment in Randomized and Nonrandomized Studies. *Journal of Educational Psychology*, *66*, 688–701.

Rubin, D. (1978). Bayesian Inference For Causal Effects. *Annals of Statistics*, *6*, 34–58.

A Aggregation

For a given magnet program, a separate lottery is conducted for each grade, and, within 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 there are $j = 1, ..., J$ 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 $a_{t,,j}$, ℓ_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 $N = \Sigma_j N_j$. The share of lottery j is then $n_j = N_j/N$. Extending our previous notation, W_{ij} equals 1 if applicant i to lottery j wins and 0 otherwise, A_{ij} equals 1 if applicant i to lottery j attends a school in the district and 0 otherwise, and M_{ij} equals 1 if applicant i to lottery j attends magnet school j and 0 otherwise..

Let $w = \Sigma_j n_j w_j$, $a_t = \Sigma_j n_j a_{t,j}$, $\ell = \Sigma_j n_j \ell_j$, $s_m = \Sigma_j n_j s_{m,j}$, $s_n = \Sigma_j n_j s_{n,j}$, and $r = \Sigma_j n_j r_j$. Thus, w, a_t, ℓ, 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 $a_{t,,j}$, ℓ_j, r_j, $s_{m,j}$, and $s_{n,j}$ for all j. The n_j are known and non-random. Hence, w, a_t, ℓ, r, s_m, and s_n are identified. We therefore focus on estimation and inference at the program level. Consider the following:

$$\frac{1}{N} \sum_{j=1}^{J} \sum_{i=1}^{N_j} \frac{W_{ij}(1 - M_{ij})A_{ij}}{w_j} \to \sum_{j=1}^{J} n_j s_{n,j} = s_n \tag{20}$$

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

$$
\begin{aligned}
&\frac{1}{N}\sum_{i=1}^{N}\frac{W_{i1}}{n_1} - w_1 \\
&. \\
&. \\
&\frac{1}{N}\sum_{i=1}^{N}\frac{W_{iJ}}{n_J} - w_J \\
&\frac{1}{N}\sum_{j=1}^{J}\sum_{i=1}^{N_j}\frac{W_{ij}M_{ij}A_{ij}}{w_j} - (r + s_m + a_t) \qquad (21)\\
&\frac{1}{N}\sum_{j=1}^{J}\sum_{i=1}^{N_j}\frac{W_{ij}(1-M_{ij})A_{ij}}{w_j} - \ s_n \\
&\frac{1}{N}\sum_{j=1}^{J}\sum_{i=1}^{N_j}\frac{W_{ij}(1-M_{ij})(1-A_{ij})}{w_j} - l \\
&\frac{1}{N}\sum_{j=1}^{J}\sum_{i=1}^{N_j}\frac{(1-W_{ij})M_{ij}A_{ij}}{(1-w_j)} - \ a_t \\
&\frac{1}{N}\sum_{j=1}^{J}\sum_{i=1}^{N_j}\frac{(1-W_{ij})(1-M_{ij})A_{ij}}{(1-w_j)} - (s_n + s_m)
\end{aligned}
$$

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

$$\frac{1}{N_j}\sum_{i=1}^{N_j} T_i W_{ji} A_{ji} M_{ji} \rightarrow w_j \left\{ r_j E[T_{1j}|R_j = 1] + s_{mj} E[T_{1j}|S_{m,j} = 1] \right\} \qquad (22)$$

Using the same logic above and pooling over lotteries implies that:

$$\frac{1}{N}\sum_{j=1}^{J}\frac{1}{w_j}\sum_{i=1}^{N_j} T_i W_{ji} A_{ji} M_{ji} \rightarrow \sum_{j=1}^{J} n_j \left\{ r_j E[T_{1j}|R_j = 1] + s_{mj} E[T_{1j}|S_{m,j} = 1] \right\} \qquad (23)$$

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

$$\sum_{j=1}^{J} n_j r_j E[T_{1j}|R_j = 1] \leq \sum_{j=1}^{J} n_j r_j U \ = \ rU \qquad (24)$$

Combining equations (23) and (24) and normalizing by s_m, we have

$$\frac{1}{s_m}\sum_{j=1}^{J} n_j s_{mj} E[T_{1j}|S_{m,j}=1] \;\; \geq \;\; \frac{1}{s_m}\left[\frac{1}{N}\sum_{j=1}^{J}\frac{1}{w_j}\sum_{i=1}^{N_j} T_i W_{ji} A_{ji} M_{ji} \; - r\,U\right] \quad (25)$$

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 $L \leq E[T_{1j}|R_j = 1]\ \forall j$, yields an upper bound for the weighted treatment effect.