Evaluating Education Programs That Have Lotteried Admission and Selective Attrition*

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

August 16, 2012

*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 a coeditor of the journal, three anonymous referees, 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 participants at numerous conferences and seminars. 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 of Education Sciences (IES R305A070117 and R305D090016).

Abstract

We study the effectiveness of selective education programs in an urban school district that ration excess demand by admission lotteries. We focus on oversubscribed magnet programs and show that these programs help the district to attract and retain students that come from higher incomes than students that stay in the district regardless of the outcome of the lottery. As a consequence, differential attrition arises since students that lose the lottery are more likely to pursue educational options outside the school district than students that win the lottery. When students leave the district, important outcome variables are often not observed, which creates a missing data problem. The treatment effects are, therefore, not point identified. We exploit known quantiles of the outcome distribution to construct informative bounds on treatment effects. The bound estimates demonstrate that magnet programs offered by the district improve behavioral outcomes such as offenses, timeliness, and attendance, but have no significant effect on achievement.

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

1 Introduction

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

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. Educational outcomes are often not observed for students that 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.[1]

Lottery-based admission can be viewed as an experimental design with multiple sources of non-compliance that arise from parental or student decisions. 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

[1]Selective attrition may also arise when lottery winners that initially participate in the program drop out because they experience unfavorable outcomes.

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 latent household types. Our approach builds on Angrist, Imbens, and Rubin (1996) and adds two additional latent types to the framework to deal with multiple sources of non-compliance. 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 percent of applicants to magnet programs that serve elementary school students are "at risk." That suggests that 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, i.e. 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 in the lottery. As a consequence, we cannot point-identify the causal effect of magnet programs on achievement or other behavioral outcomes.[2] We, therefore, focus the remainder of the analysis on estimating informative bounds on these treatment effects. One prominent approach that is

[2]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.

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 non-random 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 district wide distribution of offenses, suspension days, tardies, and absences.[3] 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 only conducted in grades 5, 8, and 11 during most of our sample period.

The rest of the paper is organized as follows. Section 2 provides a brief review of the literature. Section 3 discusses identification and estimation of treatment effects when program participation is partially determined by lotteries and selective attrition cannot be ignored. Section 4 provides some institutional background for our application and discusses our main data sources. Section 5 reports the empirical findings of our paper. Finally, we offer some conclusions and discuss the policy implications of our work in Section 6.

[3]We also implement Lee's (2009) approach which uses sample trimming rules to construct informative bounds. Note that this approach is closely related to Zhang and Rubin (2003).

2 Literature Review

Our paper is related to a growing literature that evaluates educational programs using lottery based estimators.[4] 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 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 paper.[5]

Currently, there are 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.[6] In some applications,

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

[5] 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.

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

there exists an exogenous variable that affects the selection, but not the outcome equation. These type of exclusion restrictions can be used in both parametric and semi-parametric estimation techniques.[7] 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 general framework for dealing with non-random 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. [8]

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.[9] Our approach also relies on latent types and builds on Angrist, Imbens, and Rubin (1996) to account for the multiple sources of non-compliance.

[7]Some well-known examples are Heckman (1990), Ahun and J.Powell (1993) and Das, Newey, and Vella (2003).

[8]There are two additional related papers that use bounding methods. Dinardo, McCrary, and Sanbonmatsu (2006) develops a bounding method that requires an instrument for attrition. Blundell, Gosling, 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.

[9]For a recent application see Barnard, Frangakis, Hill, and Rubin (2003), who study the effect of school choice on test scores, or Zhang, Rubin, and Mealli (2009) evaluate the Job Corps training program.

3 Identification and Estimation

3.1 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.[10] 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 district and 0 otherwise.

To model compliance with the intended treatment, we define five latent types to classify households into compliers and non-compliers.[11]

Defintion 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*

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

[11] Appendix B discusses the assumptions needed to to derive these five types from the 16 possible types.

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.*[12]

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 ob-

[12] 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.

served 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 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 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:

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

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.[13] We will later assume that T is not observed for students that do not attend a public school within the district, i.e. T_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

[13] 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.

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.[14]

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

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.

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" 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.

3.2 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.

[14]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	$(1 - w)\ (r + l)$	$\frac{r\mu_r + l\mu_l}{r + l}$

Proposition 1 *The parameters (w, r, s_n, s_m, l, a_t) 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):

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

Q.E.D.

There is one over-identifying 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 SES 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.

3.3 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. 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.[15] Conditioning on lottery outcomes, there are ten conditional expectations. Three of these pertain to

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

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
$W = 1$	$E[T_1 \| S_m = 1]$	$E[T_0 \| S_n = 1]$	$E[T_1 \| R = 1]$	$E[T_2 \| L = 1]$	$E[T_1 \| A_t = 1]$
$W = 0$	$E[T_0 \| S_m = 1]$	$E[T_0 \| S_n = 1]$	$E[T_2 \| R = 1]$	$E[T_2 \| L = 1]$	$E[T_1 \| A_t = 1]$

Note 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}.

This treatment effect is of interest to policy makers since the "complying stayers" account for the majority of students that attend magnet schools at any point of time. Our estimates suggest that 60 to 70 percent of all the students in our sample of applicants to magnet programs, and approximately 70 to 80 percent of all attending students, fall into that category. The school district and policy makers are obviously interested in finding out whether the magnet programs improve outcomes for the majority of students that 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 paper 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

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

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 $E[T_1|R=1]$.

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.[16] The lack of point identification arises from the "at risk"

[16]Recall that if $r=l=0$ our research design simplifies to the one considered in Angrist et al. (1996).

households which cause the selective attrition problem. Selective attrition is only a problem if “at risk” households have different mean outcomes than compliers.[17]

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 $E[T_1|R=1]$ 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:

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

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. We therefore define our lower bound as the parameter value for which equation (18) holds with equality.

[17] 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. Q.E.D.

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 upon them. It is easy to see that for any value between the lower and upper bound, 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 paper 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 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 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 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.

3.4 A GMM Estimator

Let θ denote the parameter vector which includes the fraction of the latent types, the means of the characteristics, and the lower and upper bounds of the various treatment effects. 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. 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:

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

and note that $E[\frac{1}{N}\sum_{i=1}^{N} f_1(A_i, M_i, W_i|\theta_0)] = 0$, where θ_0 denotes the true parameter value. Similarly we can estimate the mean characteristics of each type. Define:

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

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

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

and we have $E[\frac{1}{N}\sum_{i=1}^{N} f_3(A_i, M_i, W_i, T_i|\theta_0)] = 0$. 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 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 over-identifying restriction. We implement our GMM estimator using a standard two step approach, where the optimal weighting matrix is estimated based on a consistent first step estimator. Standard errors can be computed using the standard result for optimally weighted GMM estimators.[18]

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 set with fixed probability.[19] In our application, the identified set can be described by a set of closed intervals. As a consequence, we only report 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 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.

---

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

[19]For a more detailed discussion, see, for example, Imbens and Manski (2004) and Chernozhukov, Hong, and Tamer (2007).

4 Data

We focus 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 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 socio-economic 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 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.

Many of the magnet schools in our district are vocational in nature. These programs are always undersubscribed, and 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 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.[20] 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. 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

[20]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.

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)
Subsidized lunch	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

Standard deviations are reported in parentheses.

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. To be eligible for free lunch, households must have income below 130 percent of the poverty line. Reduced lunch eligibility requires income below 185 percent of the poverty line. The variable denoted FRL is then the indicator for free or reduced lunch status.[21] 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 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

[21]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.

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 one suspension day and 96% show three days or less of suspension. There is not any explicit description of the infraction, nor is there any clarification as to whether or not 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 one day of suspension, we believe the events we call offenses in the data files are Level 2 infractions. In other words, these are highly disruptive and definitely non-conducive 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 sub-samples that we also consider in estimation.[22] 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. We estimate a separate regression for each observable. The coefficient β_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. β_1 is not significant for any tested variable

[22]We imputed absences and tardies for a small sample of students. Also note that behavioral variables are not observed for students that apply from non-public 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.

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)
Subsidized Lunch	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)

Estimated standard errors are reported in parentheses.

at 10%. The second through fourth columns contain the other subsamples of interest. We find that the estimates of β_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 0.933 for the full sample, 0.704 for the elementary schools, 0.816 for the middle schools, and 0.522 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 sub-samples 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 below.

Table 5: Lotteries by Race

Level	Black Wins	Black Losses	Other Wins	Other Losses	Black WP	Other WP
Elem	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%

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

5 Empirical Results

5.1 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 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 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 6 and 7 report the point estimates and estimated standard errors for each of the three samples.

Table 6: 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.

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.[23] We find that there are significant efficiency gains in the estimates of two

[23]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.

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. Our framework does not generate many over-identifying restrictions. For each household characteristic we add, we obtain one over-identifying restriction. This is the case because 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 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 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 on free or reduced lunch programs than households that

Table 7: 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.

are stayers. Moreover, they come from better educated neighborhoods.[24] 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.[25]

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 rates of 52 percent and 53 percent respectively while the latter have an average win rate of 77 percent. 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 is also higher among elementary school students. These findings highlight the fact that,

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

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

among the magnet school applicants, the market for elementary school education is more competitive than the market for high school education.

5.2 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. 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 prior test scores. For the high school sample, the reduction is of similar magnitude.[26] Ultimately, we have test score outcomes for 213 middle school students (8th grade exam) and 203 high school students (11th 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

[26]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

	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.

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.

We next turn our attention to behavioral outcomes measured one year after the lotteries were conducted.[27] 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 signif-

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

Table 9: 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	-0.26	-0.26	-0.49	-0.45	-0.47
	(0.09)	(0.09)	(0.09)	(0.15)	(0.15)	(0.14)
MS	-0.62	-0.48	-0.66	-0.22	0.00	-0.56
	(0.36)	(0.36)	(0.35)	(1.17)	(1.18)	(0.77)
HS	-0.03	0.28	0.20	-0.47	0.14	0.03
	(0.34)	(0.39)	(0.31)	(0.87)	(0.93)	(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	0.98	-1.70	-0.95	0.52	-0.98
	(0.90)	(1.24)	(0.77)	(0.73)	(0.87)	(0.59)
MS	1.98	4.16	1.82	3.04	4.97	2.32
	(1.60)	(2.02)	(1.36)	(1.82)	(2.07)	(2.07)
HS	-8.64	-5.35	-7.77	-7.90	-6.61	-9.41
	(3.32)	(3.60)	(2.55)	(2.78)	(2.87)	(2.45)

Estimated standard errors are reported in parentheses.

icant 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:

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

where

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

Now if $r = 0$, then

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

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

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 HS 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.

5.3 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 state-wide 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 percentiles of the district-wide measures are 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 3 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 over-identifying conditions and implement an exactly identified estimator. For example, using the middle school sample, our estimates with (without) over-identifying conditions for the upper bound on offenses is -0.62 (-0.64), suspension days -0.22 (-0.23), absences 1.98 (1.61), and tardies 3.04 (3.02). We also implemented the over-identified estimator for non-optimal weighting matrices such as the identify matrix and again found insignificant differences. We thus conclude

Table 10: Results for Different Quantiles

	90 %	95 %	99 %
MS Math	129.15	91.08	35.66
	(177.11)	(183.69)	(198.69)
MS Read	37.52	3.68	-59.35
	(167.98)	(172.69)	(188.40)
HS Math	-8.37	-24.22	-53.31
	(131.69)	(148.00)	(179.01)
HS Read	-9.68	-25.09	-53.10
	(121.37)	(136.24)	(164.81)
ES absent	0.58	0.98	0.98
	(1.19)	(1.24)	(1.24)
MS absent	3.97	4.16	4.16
	(1.97)	(2.02)	(2.02)
HS absent	-5.92	-5.54	-5.35
	(3.50)	(3.57)	(3.60)

Estimated standard errors are reported in parentheses.

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. 16 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. Using the full sample, our estimate for the upper bound on offenses is -0.62 with an associated standard error of 0.36. Using the sub-sample with low win rates, we obtain an estimate of -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 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.[28] As we detail in the appendix, 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.[29] Hence, the comparison in Table 11 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 11, we have confirmed that the results from our weighted estimator are similar to those when we

[28] 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.

[29] 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.

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.[30] 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 paper 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. 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 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.

[30] Details are available on request.

Table 11: Comparison with Lee's 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] $\hat{p} = 0.426$

6 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 which 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 state wide achievement tests as outcome measures. As a consequence, they have access to the same outcome measure for students that decided to stay in the district and that 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 that remains 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

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.

Ahun, H. and J.Powell (1993). Semiparametric Estiamtion of Censored Selection Models with a Nonparamteric Selection Mechanism. *Journal of Econometrics*, *58*, 3–29.

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., 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.

Barnard, J., Frangakis, C., Hill, J., and Rubin, D. (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, 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., Gosling, R., 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.

Chernozhukov, V., Hong, H., and Tamer, E. (2007). Estimation and Confidence Regions for Parameter Sets in Econometric Models. *Econometrica*, *75 (5)*, 12431284.

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.

Das, M., Newey, W., and Vella, F. (2003). Nonparamteric Estiamtion of Sample Selection Models. *Review of Economic Studies*, *70*, 33–58.

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

Frangakis, C. and Rubin, D. (2002). Principal Stratification in Causal Inference. *Biometrica*, *58*, 21–9.

Frolich, M. and Lechner, M. (2010). Combining Matching and Nonparametric IV Estimation: Theory and an Application to the Evaluation of Active Labor Market Policies. Working Paper.

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. (1974). Shadow Prices, Market Wages and Labor Supply. *Econometrica*, *42 (4)*, 679–694.

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. (1990). Varieties of Selection Bias. *American Economic Review*, *80*, 313–18.

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. (2000). 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.

Imbens, G. and Wooldridge, J. (2007). Partial Identification. NBER Lecture Notes: What's New in Econometrics?

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

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

Manski, C. (1990). Nonparametric Bounds on Treatment Effects. *American Economic Review: Papers and Proceedings*, *80*, 319–23.

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

Neyman, J. (1923). On the Application of Probability Theory to Aggricultural Experiments: Essay on Principles. *Translated in Statistical Science*, *5*, 465–80.

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. (1976). Inference and Missing Data. *Biometrika*, *63*, 581–92.

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

Zhang, J. and Rubin, D. (2003). Estiamtion of Causal Effects via Principal Stratification When some Outcoems are Truncated by 'Death'. *Journal of Educational and Behavioral Statistics*, *28*, 353–68.

Zhang, J., Rubin, D., and Mealli, F. (2009). Likelihood-Based Analysis of Causal Effects of Job-Training Programs Using Principal Stratification. *Journal of the American Statistical Association*, *104*, 166–176.

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

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

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

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

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:

$$\sum_{j=1}^{J}n_jr_jE[T_{1j}|R_j=1]\leq\sum_{j=1}^{J}n_jr_jU\ =\ rU \qquad (24)$$

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.

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 Frolich and Lechner (2010). They 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 is in the same vein as Frolich and Lechner's estimates for the compliers.

B Latent Types

Defining the potential outcomes $A(W)$ and $M(W)$, the 16 possible types are given in Table 12. 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*

Table 12: Potential Types

$\star$	1.	$A(0) = 1, M(0) = 0$ and $A(1) = 1, M(1) = 0$	never taker (S_n)
$\star$	2.	$A(0) = 1, M(0) = 1$ and $A(1) = 1, M(1) = 1$	always taker (A)
$\star$	3.	$A(0) = 1, M(0) = 0$ and $A(1) = 1, M(1) = 1$	complier (S_m)
	4.	$A(0) = 1, M(0) = 1$ and $A(1) = 1, M(1) = 0$	defier
$\star$	5.	$A(0) = 0, M(0) = 0$ and $A(1) = 0, M(1) = 0$	never taker (L)
	6.	$A(0) = 0, M(0) = 1$ and $A(1) = 0, M(1) = 1$	always taker
	7.	$A(0) = 0, M(0) = 0$ and $A(1) = 0, M(1) = 1$	complier
	8.	$A(0) = 0, M(0) = 1$ and $A(1) = 0, M(1) = 0$	defier
	9.	$A(0) = 0, M(0) = 0$ and $A(1) = 1, M(1) = 0$	complier w.r. to moving, never taker
	10.	$A(0) = 0, M(0) = 1$ and $A(1) = 1, M(1) = 1$	complier w.r. to moving, always taker
$\star$	11.	$A(0) = 0, M(0) = 0$ and $A(1) = 1, M(1) = 1$	complier w.r. to moving, complier (R)
	12.	$A(0) = 0, M(0) = 1$ and $A(1) = 1, M(1) = 0$	complier w.r. to moving, defier
	13.	$A(0) = 1, M(0) = 0$ and $A(1) = 0, M(1) = 0$	defier w.r. to moving, never taker
	14.	$A(0) = 1, M(0) = 1$ and $A(1) = 0, M(1) = 1$	defier w.r. to moving, always taker
	15.	$A(0) = 1, M(0) = 0$ and $A(1) = 0, M(1) = 1$	defier w.r. to moving, complier
	16.	$A(0) = 1, M(0) = 1$ and $A(1) = 0, M(1) = 0$	defier w.r. to moving, defier

the lottery but would stay if they lose, i.e. no defiers with respect to moving.[31]

b) There are no individuals who do not attend a non-magnet district school if they lose the lottery but attend a non-magnet district school if they win the lottery.

c) There are no defiers (monotonicity assumption).

Assumption (a) eliminates types 13 through 16 in the list above. Assumption (b) eliminates 9. Assumption (c) eliminates 4. This leaves our five types, denoted above with * in Table 12.

[31] This is the same monotonicity assumption used, for example, in Lee (2009).