Family Networks and School Choice

Andrew Dustan*
Department of Economics
Vanderbilt University

March 5, 2018

Abstract

This paper uses variation in school assignments generated by Mexico City's public high school choice mechanism to document a strong causal influence of older siblings on their younger siblings' school choice behavior. The effects of older sibling admission on the probabilities of choosing both the sibling's school and distinct but observably similar schools are large and positive, even when siblings are too far apart in age to attend school together. The evidence is more consistent with information transmission and path dependence channels than cost, convenience, sibling competition, or parental pressure. Sibling-induced changes in stated preferences affect admissions outcomes, including assignment to elite schools. The results imply significant externalities from policies such as affirmative action that alter the distribution of school assignments, as well as providing insight into the role of family networks in decision-making regarding educational investments.

JEL Codes: I20, I25, O15

*E-mail: andrew.dustan@vanderbilt.edu. I thank Alain de Janvry, Elisabeth Sadoulet, Fred Finan, Ben Faber, Jeremy Magruder, and Ted Miguel for helpful comments. Participants at various UC Berkeley seminars, the Northeast Universities Development Consortium Conference, the Pacific Conference for Development, Fourth SOLE/EALE World Meetings, and several other seminars provided useful feedback. Roberto Peña Reséndiz supplied the COMIPEMS data on behalf of the Secretariat of Public Education and offered excellent insight into the school selection process. Rafael de Hoyos and the advisors at the Subsecretariat of Upper Secondary Education provided invaluable assistance with the data in early stages of the project. This research was funded in part by grants from the Institute of Business & Economic Research and the Weiss Family Fellowship through the Center for Effective Global Action. This paper previously circulated under the title "Peer Networks and School Choice under Incomplete Information."

1 Introduction

School choice systems abound in developing and middle-income countries. These systems, which are especially prevalent at the secondary and post-secondary levels, often take the form of a centralized public school choice and assignment mechanism, but voucher regimes allowing families to freely select private schools are becoming common as well.[1] The choice of school is potentially a high-stakes decision, to the extent that the school attended is important for academic and later-life outcomes and that transferring between schools outside of well-defined transition points is costly or even impossible.

The determinants of school choice behavior have thus been the subject of intense policy interest and academic investigation in both developing and developed countries, in particular when students and families face an environment with many feasible alternatives and incomplete information. Ajayi (2013), for example, documents the importance of socioeconomic status as a determinant of public high school choice in Ghana. In the United States, Hoxby and Avery (2012) find that many low-income high-achievers do not apply to high-quality universities that would likely admit them with generous aid, while Hastings and Weinstein (2008) highlight the importance of information about school quality in choosing public primary and secondary schools. It is clear that factors beyond cost, distance, and objective measures of quality play important roles in shaping revealed preferences over schools, yet these determinants are often difficult to detect and understand using administrative data on choices or attendance.

This paper is the first to document the importance of one such causal determinant of students' school choices: family networks. I show, using detailed student-level microdata from fourteen years of Mexico City's public high school choice system, that students' choices

[1] Examples of centralized public school choice systems at the secondary level include China (Lai, Sadoulet, and de Janvry 2009, 2011; Zhang 2016), Ghana (Ajayi 2013), Kenya (Lucas and Mbiti 2012, 2014), Malawi (Hoop 2011), Romania (Pop-Eleches and Urquiola 2013), Trinidad and Tobago (Jackson 2010), and Mexico (Dustan, de Janvry, and Sadoulet 2017; Estrada and Gignoux 2017). Chile's extensive voucher system is documented in Mizala and Urquiola (2013) and Hsieh and Urquiola (2006), Angrist et al. (2002) study a Colombian voucher program, and Muralidharan and Sundararaman (2015) test a voucher system in India.

and eventual admissions outcomes are impacted strongly by where their older siblings went to school. The assignment mechanism generates exogenous variation in the school assignment of older siblings, overcoming the well-known problems with causally interpreting correlated behavior in peer groups (Manski 1993, 1995). This unified choice-based allocation system determines assignment priority solely on the basis of an exam score. A sharp regression discontinuity (RD) design is thus employed: given a group of older siblings who want to attend a certain school, some score barely high enough to be assigned there while others score barely too low and must attend another school. This variation in assignment near the admission cutoff is used to identify the effects of older sibling assignment in both reduced form and discrete choice models of school choice.

The richness of the demographic and outcome data, as well as the wide variety of schools, allow me to explore the possible channels through which this sibling effect operates. This is important for understanding both the likely external validity and the policy implications of the results. The effect is not driven by the obvious explanation that it is convenient or beneficial for the family to have two children attending the same school. Older sibling admission to a particular high school increases the revealed preference for that school even when the siblings are far enough apart in age that the older sibling no longer attends high school. Nor is the effect due to a cost of changing parental commuting arrangements, as the results persist after accounting for admission's effect on geographical preferences.

The effect cannot be explained by sibling competition or changing parental expectations for the younger child regarding the prestige or competitiveness of chosen schools. Sibling admission affects preferences even when the counterfactual is admission to a similarly-competitive school, whether measured by the school's admission cutoff score or belonging to an elite school system. This is true even among same-gender sibling pairs and those with similar levels of achievement, subgroups where sibling competition is often expected to be stronger. This is in contrast with the quite different empirical context in Joensen and Nielsen (2018), who argue for the role of competition in determining high school *course* selection within Danish high schools.

The results point most strongly to the importance of information-sharing and of some form of path dependence within the family. Firstborn children's assignment outcomes have the strongest impact on younger sibling choices, suggesting that students and their parents are unsure of which school (or type of school) is the best fit for them and thus rely on the precedent set by the firstborn. An important nuance here is that the choice effects are not limited to the older sibling's exact school. Demand also increases substantially for other schools belonging to the same subsystem, within which individual schools throughout the city share many attributes such as curriculum and vocational orientation. While not inconsistent with a pure imitation channel, this finding raises the possibility that students learn about both specific schools and *types* of schools from their older siblings, leading to increased demand if students prefer "familiar" schools or if they obtain school- and subsystem-specific knowledge that increases the returns to attending there. The online appendix presents a model in which, even when student beliefs about school attributes are unbiased, risk aversion implies that the expected effect of new information about a school is an increase in demand for it. A final piece of evidence favors the information-sharing channel, differentiated from a more generic path dependence explanation: demand for a school increases much more when the older sibling experiences a positive academic outcome there (graduation), although this relationship cannot be interpreted strictly as a causal one. Working against an explanation of more naïve salience effects, I find that admission effects are as strong for schools very close to home as they are for distant ones.

Two additional results illustrate the importance of siblings in the choice process. First, the altered school choices induced by sibling assignment lead to significant changes in the assignment outcomes of the younger sibling. Students are more likely to be admitted to their older siblings' assigned schools, as well as to a school within the same subsystem. Among the students least likely to apply to the set of Mexico City's elite public high schools, sibling assignment to an elite school substantially increases the probability of both elite application and assignment. This suggests a role for information or other network effects in encouraging qualified students to apply to high-quality schools. Second, integrating the RD design into a simple discrete choice model of school choice, I find that the average marginal willingness to travel due to older sibling admission is large, in excess of six kilometers round-trip. The discrete choice model is also crucial for testing the possibility that siblings affect demand for geographical areas rather than specific schools, as discussed above.

Although the precise policy implications depend to a degree on the mechanisms through which siblings affect choice, these findings are relevant for multiple reasons. First, they indicate that interventions aimed at changing where students go to school are likely to have substantial spillover effects on younger siblings (and perhaps other members of the targeted students' networks). For example, even a temporary affirmative action policy would not only have the direct impact of changing assignment probabilities, but also an indirect impact on the demand of younger siblings due to changes in realized assignment outcomes. This would imply that the full effect of affirmative action policies may take some time to materialize. Second, if indeed students are relying on their siblings due to a lack of complete information about the schools in their choice set, then the large magnitude of the sibling effect suggests that there is an opportunity for schools or school systems to make more information available directly to students and families. The optimal type of information to provide depends on whether the lack of information is about general school characteristics and quality, as in Hastings and Weinstein (2008), or about student-school match, as in Hoxby and Turner (2013). Finally, the findings give insight into the school choice deliberation process and highlight the fact that parents are not the sole (or perhaps even the most important) members of the family network that affect choice.

The possibility that siblings' educational histories affect choice has received recent attention in the economics literature: Goodman et al. (2015) find that in the United States, siblings' college enrollment decisions are correlated. They conclude with a call for more investigation into the possibility that this relationship is causal rather than due to family-level unobservable characteristics. Hoxby and Avery (2012) make the more general point that college choice may depend on the institutions attended by the student's broader network of older peers and mentors, although the question of the network's empirical importance is left open. Researchers in education and sociology have examined the issue as well, again in the context of college choice in the United States. Kaczynski (2011) finds qualitative evidence for several of the channels discussed above in her interviews of twelve first-year college students in the United States, in particular that increased knowledge about siblings' colleges increased their probability of enrolling there. Ceja (2006) and Elias McAllister (2012) also find interview-based evidence that older siblings are a vitally important source of information for Mexican-American students as they apply to colleges. Ball and Vincent (1998) find that, for primary schools in the United Kingdom, parents use their social networks (the "grapevine") to obtain specific, detailed information about schools and their likely fit for their own children, highlighting the potential for the schooling experiences of one child (whether a sibling or not) to influence the choices for another.

The literature in economics studying the determinants of school choice is too vast to review thoroughly here. The important roles of distance from home and observed school quality have been documented extensively: see Ajayi (2013), Chumacero, Gomez, and Paredes (2011), and Hastings, Kane, and Staiger (2009) for examples in Ghana, Chile, and the United States, respectively. Perceived costs are important as well, especially in higher education (Hoxby and Avery 2012; Hoxby and Turner 2013), as is socioeconomic composition, illustrated by Burgess et al. (2014) in the United Kingdom and Hastings, Kane, and Staiger (2009). The importance of information for choice, posited to be an important channel in the present paper, has been documented as well. Hastings and Weinstein (2008), for example, demonstrate that providing information on test score aggregates to low-income families in the United States increased the likelihood of choosing high-performing schools. Related studies by Koning and Wiel (2013) in the Netherlands and Friesen et al. (2012) in Canada come to similar conclusions, while Mizala and Urquiola (2013) find no effect of publishing a quality measure in Chile. The importance of information originating from students' networks for school choice has not been quantified empirically, to my knowledge.

The paper proceeds as follows. Section 2 describes the public high school choice system in Mexico City, showing that it provides a good context in which to empirically examine school choice under incomplete information. Section 3 explains the data. Section 4 gives the reduced form RD method and results, while Section 5 lays out a simple discrete choice model and corresponding results. Section 6 concludes.

2 High school choice in Mexico City

This section explains Mexico City's public high school choice system. In addition to providing context for the empirical exercise, this section describes the assignment mechanism that is the basis for exogenous school assignment of peers and highlights some features of the system that induce students to reveal their true school preferences.

2.1 The COMIPEMS assignment mechanism

Prior to 1996, the nine major public high school subsystems in Mexico City controlled their own independent admissions processes.[2] Students applied to schools in one or more of these subsystems, waited to learn where they had been admitted, and then withdrew from all schools except their most-preferred one. In an effort to increase both the efficiency and transparency of this process, the subsystems formed the Comisión Metropolitana de Instituciones Públicas de Educación Media Superior (COMIPEMS) in 1996. Each year, COMIPEMS runs a unified, competitive admissions process that assigns students across Mexico City's public high schools on the basis of students' preferences and the results of a standardized exam.

The COMIPEMS admissions process is as follows.[3] In late January, students in ninth grade—the final year of middle school—receive informational materials about the admissions process. These materials include a list of all of their "educational options," which in most

[2]The discussion in this section draws on Dustan, de Janvry, and Sadoulet (2017).
[3]The timing of each step is given for the 2011 competition.

cases are schools but can also be specific tracks within schools, such as vocational education tracks in a technical school. Students then fill out a registration form, a demographic survey, and a list of up to 20 educational options, ranked in order of their preference. These forms must be submitted in late February or early March, depending on the student's family name. In June of that year, students take a standardized exam consisting of 128 multiple-choice questions, covering both subject-specific material from the public school curriculum and more general mathematical reasoning and language areas.

In July, the assignment process is carried out by the Centro Nacional de Evaluación para la Educación Superior (CENEVAL).[4] First, the school subsystems report the maximum number of seats available to incoming students. Second, all students who did not successfully complete middle school or scored below 31 of 128 points are discarded. Third, all remaining students are ordered by their exam score, from highest to lowest. Fourth, a computer program proceeds sequentially down the ranked list of students, assigning each student to his highest-ranked option that still has a seat remaining.[5] The process continues until all students are assigned, with the exception of students who scored too low to enter any of their listed options. Later in July, the assignment results are disseminated to students. Through 2011 this primarily happened in the form of a printed gazette sold at newsstands, although a system that sends personalized results via text message has become more popular over time. At that time, students who were eligible for assignment but were left unassigned during the computerized process because they scored too low for any of their choices may choose a schooling option from those with seats remaining.

[4] CENEVAL is independent of COMIPEMS and its constituent school subsystems. This process is carried out by computer in the presence of representatives from all subsystems and external auditors from a large international accountancy firm.

[5] In the instance that two or more students have the same score and highest-ranked available option, but there are fewer remaining seats than the number of tied students, the assignment process pauses and representatives from the corresponding subsystem must decide to either admit all tied students or none of them.

2.2 Student decision-making under the COMIPEMS mechanism

Students have some information about basic school attributes when they choose schools, but this information is generic rather than individually tailored. The subsystem membership of each school is known with certainty, and each subsystem has a well-formed public perception. There are two "elite," university-affiliated subsystems: the Universidad Nacional Autónoma de México (UNAM) and the more technically-focused Instituto Politécnico Nacional (IPN). These are believed to be competitive, relatively rigorous, prestigious high schools, and as such they are highly demanded. The distribution of cutoff scores—the exam score of the student admitted to the school's final seat—for schools that filled their capacity in 2011 is given in Figure 1, showing that elite schools have significantly higher cutoffs than most non-elite schools. Students assigned to elite schools in 2011 had entrance exam scores more than 1.6 standard deviations higher than those assigned to non-elite schools, middle school grade point averages 0.7 higher, and parents with 2.3 more years of education. Dustan, de Janvry, and Sadoulet (2017) show that assignment to an IPN school has positive effects on students' standardized exam scores, although students with low middle school GPAs are more likely to drop out from them, while Estrada and Gignoux (2017) find a positive effect of IPN assignment on students' expected returns to higher education.

Non-elite subsystems include those with traditional academic curricula and technical subsystems providing academic coursework combined with vocational training for careers such as auto repair and bookkeeping. Even within a subsystem, some official information about school-level academic quality is available. Past cutoff scores for each educational option have been available on the COMIPEMS web site since 2005, and this site is actually browsed by many students because it allows them to easily complete most of the registration process online. Cutoff score and the mean score of admitted students are almost perfectly correlated, so students have access to an excellent proxy for mean peer ability if they seek it out. The combination of subsystem reputations and information about peer quality ensures that students are at least somewhat informed about general school attributes, though they may lack more specific details that affect the idiosyncratic match between the student and school.

According to the 2011 Mexican high school census, the median tuition of COMIPEMS-participating high schools was about $110 USD per year. Uniforms, textbooks, and other materials were estimated to cost an additional $150 USD. Importantly, public schools do not offer discounts to students who have siblings in the same school or subsystem, so there is no direct financial benefit to choosing the older sibling's school.

When asked about the choice process, most administrators and students claim that it is the student himself who decides on the schools he lists and the order in which they are listed. Parents are also involved to varying degrees in the deliberation process, offering another channel through which older sibling admission outcomes may affect school choices. Notably, though, the parent's signature is not required on the choice form: a school official can provide it. Students often construct their rankings in the following way, similar to how United States students choose colleges (see Hoxby and Avery (2012), for example). First, they decide whether they would like to attend a high school in either or both of the elite subsystems. If a student decides to apply within either or both subsystems, he lists some number of elite schools as his top choices. There are 30 elite schools (16 IPN and 14 UNAM), meaning that even within an elite subsystem, students face a wide variety of options. Following the elite schools, if any, he lists various non-elite schools (from about 600 options in most years), which offer a better chance of admission.[6]

Two aspects of the COMIPEMS assignment mechanism make the student's ranking quite informative about true preferences. First, the mechanism is equivalent to the deferred acceptance algorithm proposed by Gale and Shapley (1962), so it induces truth-telling by students.[7] Under such mechanisms it is never optimal to list a less-preferred school before a more-preferred school, regardless of the limit on how many options can be listed. Second,

[6] Most students live within a reasonable commuting distance of many schools, as illustrated in Appendix Figure B.1.

[7] See Dubins and Freedman (1981) and Roth (1982). This particular mechanism is referred to as a student-proposing deferred acceptance mechanism, which is discussed in Abdulkadiroglu and Sonmez (2010).

the ability to rank up to twenty options means that few students actually fill up their entire preference sheet; students generate a satisfactory choice portfolio without confronting the space constraint.[8] There is no strategic disadvantage to choosing a school at which the student has a small *ex ante* probability of admission, both because the number of options allowed is high and because the assignment algorithm does not punish students for ranking unattainable schools.

3 Data and sample construction

This section describes the Mexico City public high school admissions data and the sample construction that forms the basis for the RD design.

3.1 Data description

This paper uses administrative data compiled by COMIPEMS for fourteen admissions cycles, from 1998 to 2011. For each student who registered for the exam, the database contains basic demographic information including the student's full name, date of birth, phone number, address, and a unique middle school identifier along with the grade point average attained there; the full list of up to 20 ranked school preferences; a context survey, completed by the student, including information about parental education, family composition, and other topics; and assignment results, including the student's exam score and the school assigned during the computerized allocation process.

The analysis is limited to sibling pairs where the older sibling attended a public middle school and each sibling was taking the exam in the final year of middle school. Demographic information is used to match siblings with each other in the following way. First, potential siblings are identified if they have the same paternal and maternal family names and either 1) have the same phone number or 2) live in the same postal code and attend the same middle school. From this pool of potential matches, sibling pairs are discarded if 1) either student states that he has no siblings; 2) the students report birth orders indicating that

[8] Choosing the optimal portfolio of schools is a complex problem if listing choices is costly (e.g. time cost or opportunity cost due to a limited number of allowed choices), as mentioned by Ajayi (2013). Chade and Smith (2006) model a related portfolio choice problem and derive its solution.

they are not the closest siblings in the family;[9] 3) the students were born fewer than nine months apart or more than six years apart, the latter because it is unlikely that consecutive births more than six years apart represent a true match;[10] or 4) the older student took the exam after the younger one. If one student matches with two potential older siblings, the match based on the shared phone number is used.

The performance of this matching algorithm and its implications for the paper's results are discussed in detail in Section C of the online appendix. In brief, the algorithm finds matching younger siblings for 39% of students who are likely to be older siblings based on their reported birth order and number of siblings. This incomplete matching is expected, given that families may move between COMIPEMS cycles, siblings may attend different middle schools, and students may supply different phone numbers (including personal cell phone numbers) for a variety of reasons. It is unlikely that a large proportion of unmatched older siblings are due to the younger sibling not participating in COMIPEMS, since dropout between middle school and high school is rare in urban Mexico and COMIPEMS is the only avenue by which students can access Mexico City's public schools. While the matching is imperfect, its success rate is estimated to be unaffected by the older sibling's school assignment. This suggests that the RD design is internally valid. However, matching success varies somewhat with baseline student performance and demographic characteristics. I show that the estimated average treatment effects are highly robust to inverse probability weighting with respect to the estimated propensity to be included in the sample.

The matching process locates 540,118 sibling pairs in a population of 3,051,216 students. Columns 1 and 2 of Table 1, Panel A give a description of demographic, academic, and school choice variables for the full sample of students and for the matched younger siblings, respectively. The matched younger siblings are quite similar on average to the full sample, although the differences in means are statistically significant. The average student ranks 9

[9]This is done so that the estimated effect of older sibling of admission does not include an indirect effect through the influence on a middle sibling's behavior.

[10]Only 7% of matched siblings meeting the other criteria were more than six exam years apart.

school choices, which is similar across samples. Almost two thirds of students select a school in one of the two elite subsystems as their first option, but only one in five are admitted to one. On average, students choose a school almost 8 kilometers away as their first option, measured as a straight line from the center of the student's home postal code to the school.[11] Column 3 shows that younger sibling observations in the RD sample for a bandwidth of 10 points (to be explained in the next section) are quite similar to the full sample of younger siblings, except that they rank more schools on average and are 8 percentage points more likely to choose an elite school as their first choice. The number of observations in the RD sample reflects the inclusion of the same sibling pair multiple times when the older sibling is near the admission cutoff of multiple schools. The total number of unique younger siblings in column 3 is 288,863.

To measure whether the older sibling graduated or dropped out of high school, the COMIPEMS database is merged via national ID number (CURP) with a database from the national 12$^{\text{th}}$ grade exam, called the ENLACE Media Superior. This exam is given to students who are on track to graduate at the end of the academic year, so it is a good proxy for graduation.[12] This exam was only administered starting in the spring of 2008, and the database used in this paper contains results from 2008 to 2012, corresponding to students taking the COMIPEMS exam in 2005-2008 because high school is three years in Mexico and some students take four years to graduate. Thus the part of the analysis using this graduation measure is limited to younger siblings of these cohorts, which limits sample size. The larger and more demanded of the two elite subsystems, the UNAM, does not administer the ENLACE exam so graduation data are missing for students assigned there. This further limits the sample size when the graduation measure is used. About 54% of students assigned to a non-UNAM high school between 2005 and 2008 took the ENLACE exam.

Table 1, Panel B describes the sibling pairs in more detail. Matched siblings are, on

[11]Postal codes are very geographically specific in Mexico City. Students in the sample belong to more than 2,800 postal codes.

[12]For more details on the ENLACE and how it relates to graduation, see Dustan, de Janvry, and Sadoulet (2017).

average, 2.8 grade years apart. About half are same-sex pairs. Younger siblings only have 0.03 lower middle school GPA on average, although the variance of this difference is quite high and, in absolute value, the between-sibling GPA difference is 0.73 points. Siblings have fairly similar high school preferences: 33% of sibling pairs select the same school as their first choice. The RD sample is very similar on these dimensions.

3.2 Overview of empirical strategy and sample definition

The COMIPEMS school assignment mechanism provides exogenous variation in older siblings' school assignment because, conditional on the older sibling's ranking of schools, his assignment depends solely on his exam score. This permits the use of a sharp RD design, similar to those used in prior work investigating the academic effects of school assignment in exam-based allocation regimes.[13] The basic idea behind this design is to define, for each school, the sample of older siblings who were either marginally admitted or marginally rejected from that school, and then compare the choices and outcomes of the younger siblings in the marginally admitted and marginally rejected groups. The rest of this subsection gives the procedure for defining the "marginal" sample of older siblings for use in the RD analysis.

The assignment process results in hard cutoff scores for each educational option that filled all of its seats and thus had to reject some students; this cutoff is equal to the lowest score among all admitted students. For simplicity I will refer to an educational option as a school in this discussion. Define this cutoff as c_j for school j. (The cutoff score for a given school varies across years, but for notational simplicity in the present discussion I assume there is only one year of data.) If school k is ranked before j on student i's preference list, including if j is unlisted, we write $k \succ j$.[14] Denote the student's exam score as s_i. Then marginal students for school j are those who:

1. listed school j as a choice, such that all schools preferred to j had a higher cutoff score

[13] See, for example, Pop-Eleches and Urquiola (2013), Abdulkadiroglu, Angrist, and Pathak (2014), Dobbie and Fryer (2011), Clark (2010), Jackson (2010), Hoop (2011), and Dustan, de Janvry, and Sadoulet (2017).

[14] This notation is not meant to suggest that the student must prefer his stated preferences to those that are not listed. It merely denotes the ordering of the student's ranking for the purposes of assignment.

than j (otherwise assignment to j is impossible): $c_j < c_k, \forall k \succ j$, and

2. had a score sufficiently close to j's cutoff score to be within a given bandwidth w around the cutoff: $-w \leq s_i - c_j < w$.[15]

This marginal group includes students who were rejected from j ($s_i < c_j$) and those who scored high enough for admission ($s_i \geq c_j$). A student may belong to more than one school's sample of marginal students. Not all students scoring high enough for admission are actually assigned to j; some score sufficiently high for admission to some $k \succ j$. Figure 2 plots the probability of being assigned to the cutoff school as a function of $s_i - c_j$ and verifies that the jump in probability of assignment to the cutoff school at the cutoff score is exactly 1. This probability falls monotonically with score to the right of the cutoff, as higher-scoring students are assigned to more-preferred schools. Because the RD design identifies the "admission effect" for students at the cutoff and the probability of assignment to the cutoff school changes from 0 to 1 there, one can think of the RD coefficients as "assignment effects."

Two more restrictions are placed on the sample, not to fulfill the assumptions of RD but to ease interpretation of the admission effect. First, students who would be unassigned to any school upon missing j's cutoff by a point are omitted. Such students did not list any school with a cutoff score equal to or less than c_j. We do not know if the unassigned students later chose a school from those that did not fill up or if they did not enroll at all. Second, I also omit students who are at the margin of admission to a particular career track within a technical school and who would be assigned to another track within the same school upon missing the cutoff by a point. These students only change career track upon rejection, not school. Our focus is thus on the effect of a sibling being assigned to one school vs. another, rather than 1) getting into any school vs. going unassigned or 2) being assigned to one track vs. another within the same school.

[15] The second inequality is strict because the score variable is discrete, so this definition includes w score values too low to be admitted and w values high enough to be admitted.

4 Reduced form regression discontinuity analysis

The reduced form analysis provides clear, easily-interpreted evidence about the effects of sibling assignment on school choice and admission outcomes. Much of the logic from this analysis will be applied when estimating the discrete choice model in the next section.

4.1 Method

For all regressions in this paper, COMIPEMS exam score is centered to be 0 at the school's cutoff score, which may be different in each year t: $\tilde{s}_{ijt} \equiv s_i - c_{jt}$. Most of the analysis uses the following stacked nonparametric RD specification, which combines the information from all oversubscribed schools in order to make statements about the average effect of admission:

$$y_{ijt} = \delta admit_{ijt} + \beta_1 \tilde{s}_{ijt} + \beta_2 \left(admit_{ijt} \times \tilde{s}_{ijt} \right) + \mu_{jt} + \varepsilon_{ijt} \quad (1)$$

where y_{ijt} is the outcome of interest for student i with older sibling near the cutoff of school j in exam year t, $admit_{ijt}$ is a dummy variable for whether $\tilde{s}_{ijt} \geq 0$,[16] $\beta_1 \tilde{s}_{ijt}$ and $\beta_2 \left(admit_{ijt} \times \tilde{s}_{ijt} \right)$ are linear terms in exam score approximating the unobservables that vary with score, μ_{jt} are cutoff school-exam year fixed effects, and ε_{ijt} is an error term.[17] In our case, y_{ijt} is an outcome for the younger sibling, such as choosing school j as his first option, while the explanatory variables are from the older sibling. The parameter δ is the average treatment effect of the older sibling's admission to the cutoff school on the younger sibling's outcome. Because this effect is estimated for older siblings near an admissions cutoff, it represents the average effect of older siblings' marginal admission compared to the counterfactual of marginal rejection from the cutoff school and admission to the most-preferred school that would actually accept them.

[16] The variable $admit_{ijt}$ is equal to 1 both for students actually assigned to j and students who scored high enough to be admitted to a more-preferred school.

[17] It might be preferable to include different slope coefficients for each school or school-year, similar to Abdulkadiroglu, Angrist, and Pathak (2014) who include different functions for each school. But the large number of schools makes this infeasible, so I include only one set of first-order polynomials, as in Pop-Eleches and Urquiola (2013).

This nonparametric local linear regression model uses the edge kernel to weight observations. Bandwidth selection is performed using the procedure proposed in Imbens and Kalyanaraman (2012). Because it is unclear how well a nonparametric estimator will perform with the discrete running variable used here and because the IK bandwidth is large in some cases, I also present parametric RD estimates of all results in section D of the online appendix for fixed bandwidths of 5 (linear) and 10 (linear and quadratic). The nonparametric and parametric results are very similar.

Standard errors must account for the fact that each older sibling can contribute multiple observations if he is near two or more cutoffs (or has twin younger siblings). Lee and Card (2008) recommend that when the running variable is discrete, as is the case here, standard errors should be clustered at the level of the running variable. This often results in relatively few clusters in the present application, creating a situation where two-way clustering (sibling pair and running variable) is necessary and in which one dimension may have few clusters. Cameron and Miller (2015) note that there is currently no known ideal way to handle this situation. I address it by estimating cluster-robust standard errors with respect to the older sibling and using a $T(G-1)$ distribution for hypothesis testing, where G is the number of distinct values of the running variable in the regression.[18]

Heterogeneous impacts with respect to student and sibling pair characteristics are estimated using the following specification:

$$\begin{aligned} y_{ijt} = \delta admit_{ijt} + \beta_1 \tilde{s}_{ijt} + \beta_2 \left(admit_{ijt} \times \tilde{s}_{ijt}\right) + \mu_{jt} + \\ x_i \left\{\gamma + \alpha admit_{ijt} + \beta_3 \tilde{s}_{ijt} + \beta_4 \left(admit_{ijt} \times \tilde{s}_{ijt}\right)\right\} + \varepsilon_{ijt}, \end{aligned} \tag{2}$$

[18] This approach was found to be more conservative than using two-way clustering and using the $T(G-1)$ distribution, likely because the specification error in the local polynomial appears to be small and so any small increase in the standard error due to clustering on this dimension is more than offset by the downward bias in standard errors due to the few clusters problem. Another approach with few clusters is to use the wild-cluster bootstrap to obtain p-values for hypothesis tests (Cameron, Gelbach, and Miller 2008). It is not clear that this will work well when a second dimension of clustering is introduced. In exploring this issue, bootstrapping on the running variable while computing two-way clustered standard errors for each iteration was almost always less conservative than the approach that I chose to use here.

where x_i is the characteristic for which heterogeneous impacts are to be estimated. The parameter of interest is α, the average differential impact of older sibling admission on the outcome of the younger sibling. A specification allowing different baseline admission effects for each cutoff school is also estimated, which permits estimation of heterogeneity that is not due to the fact that students with different values of the covariate of interest are close to the cutoffs of different kinds of schools.

4.2 Average effect of older sibling admission on school choice

The RD estimates give consistent causal evidence that students are more likely to apply to a school and to rank it highly if an older sibling was assigned there. Table 2 presents the estimated effects, accompanied by graphical evidence in Figure 3. Column 1 shows that for the local linear regression specification, older sibling admission to the cutoff school increases the probability of choosing the school as first choice by 7.3 percentage points (p.p.).[19] This estimate is large compared to the corresponding 14.3% choosing the cutoff school among students missing the admission cutoff by one point. Admission increases the probability of choosing the cutoff school as any choice by 10.5 p.p., compared to 60.6% for students one point below the cutoff. Effects on the probability of choosing the school immediately below the cutoff (the school to which a student is assigned if he misses admission to the cutoff school by one point) are similar: 5.2 p.p. for first choice probability and 12.7 p.p. for choosing the school at all.

The admission effects for most schools are positive. Figure 4 shows the distribution of estimated admission coefficients, obtained by estimating the RD specification separately for each school. Panel A gives the distribution of admission effects on first choice demand for elite cutoff schools only, which have large corresponding sample sizes and thus fairly precisely estimated effects. All but two of the 30 schools have positive estimated effects of admission. Panel B shows the distribution of the effect on first non-elite choice for non-elite cutoff schools. Here, estimation error overstates the variance of the distribution substantially, such

[19] When the cutoff school has multiple career tracks, the dependent variable is equal to 1 when the younger sibling chooses any track in the school.

that the estimated effect of admission is negative for 26% of schools. To account for the estimation error, I estimate the true variance of the admission effects, following Aaronson, Barrow, and Sander (2007).[20] Performing this correction and assuming a normal distribution of admission coefficients, it is estimated that only 13% of non-elite schools have a negative admission effect.

4.3 Channel: desire to attend school with or near siblings

The results are not driven by students' (or their parents') desire to attend school with their siblings, for instance because it is convenient for commuting or attending school functions. The estimated effect of admission is very similar between siblings who are close enough in age to attend high school at the same time (two or fewer grade years apart) and siblings who are too far apart in age to attend contemporaneously (three or more years, since high school is three years in Mexico). Table 3 shows this result. The estimated difference in admission effects on first choice demand for the cutoff school between these two groups is precisely estimated and close to zero: the estimated differential for further-spaced siblings in column 1 is -0.9 p.p. with 95% confidence interval of [-1.9 p.p., 0.1 p.p.]. Column 2 adds cutoff-specific admission effects, resulting in a significant estimated effect of -1.1 p.p. Columns 3 and 4 show negative, insignificant effects on ranking the school as any choice. The differential effect on first-choice demand for the school below the cutoff, reported in columns 5 and 6, is -1.6 to -1.7 p.p. and statistically significant at the 1% level, indicating a modestly reduced effect for farther-spaced siblings. But columns 7 and 8 find no differential effect on listing the school below the cutoff as *any* choice.

Furthermore, the effect of admission is not confined to demand for the older sibling's school. Admission leads to the student ranking additional schools from the same subsystem, besides the older sibling's exact school. Table 4, accompanied by Figure 5, shows that this is the case. The sample definition here is different than in the previous analysis because it only

[20] This is done by subtracting the average estimation error from the variance of the estimated coefficients: $\mathrm{E}\left[\widehat{\boldsymbol{\delta}}'\widehat{\boldsymbol{\delta}}\right] - \mathrm{E}\left[\left(\boldsymbol{\delta} - \widehat{\boldsymbol{\delta}}\right)'\left(\boldsymbol{\delta} - \widehat{\boldsymbol{\delta}}\right)\right]$.

considers students who would leave their subsystem if marginally rejected from the cutoff school. The counterfactual to admission to the cutoff school in this case is admission to a school in a different subsystem. When the older sibling is admitted to the cutoff school, the younger sibling ranks on average 0.21 more options in the same subsystem, excluding the older sibling's school, compared to an average of 1.9 schools for students one point below the cutoff. The estimated effects are almost identical for the closely-spaced and far-apart sibling samples. Similar results are found, in columns 4 through 6, for the effect of rejection from the cutoff school on demand for schools belonging to the subsystem of the school immediately below the cutoff. The admission effect on subsystem demand and the persistence of admission effects for students far apart in age cannot be explained by a preference for attending the same school contemporaneously.

Another possible channel of older sibling admission on choice is that younger siblings and their parents prefer to go to school close to where older siblings attend or have attended in the past, for example if parents set commuting patterns based on older sibling school location. This is difficult to test in the reduced form specification. The discrete choice analysis in Section 5 will show that this proximity channel seems to operate, but explains only a negligible fraction of the total impact of older sibling admission.

4.4 Channel: sibling competition or increased parental expectations

Sibling admission effects might reflect competition: when the older sibling gets into a higher-cutoff school, the younger feels the need to compete by attempting the same. Indeed, this is the channel proposed by Joensen and Nielsen (2018) as an explanation for younger students being induced to choose more challenging math and science classes. Alternatively, parents may exert pressure on the younger child when the older gets into a better school. Three related approaches are taken to explore this possibility, each failing to find clear evidence that competition and increased parental pressure are at play.

First, if competition is an important channel, then the effect of sibling admission to the cutoff school should be larger when the school below the cutoff is relatively much less competitive. Table 5 shows the effects of older sibling admission, allowing the effect size to differ based on the difference in cutoff scores between the cutoff school and the school below the cutoff. The sample is restricted to cases where both schools are in the same subsystem, so that the comparison is for differentially competitive schools within a subsystem. More precisely, observations are divided based on whether the difference in cutoffs is greater or less than the median difference in this sample (6 points on the COMIPEMS exam, about 1/3 of a standard deviation in student-level scores). Columns 1 and 2 show quite precise estimates of differential effects on first choice demand that are very close to zero. Columns 3 and 4 find small, insignificant differential effects on listing the cutoff school as any choice students facing above-median changes in cutoff score. The first-choice and any-choice results together provide no evidence for a competition channel with respect to within-subsystem school competitiveness.

Second, since the most salient quality difference between schools comes from belonging to one of the two elite subsystems, sibling competition should imply larger effects when marginal rejection from one elite subsystem results in the student being assigned to a school in a non-elite subsystem (as opposed to earning a spot in the other elite subsystem). Columns 5 through 8 of Table 5 present evidence on this comparison, limiting the sample to students at the cutoff of admission to one of the elite subsystems and allowing effects to differ with respect to whether the school below the cutoff is in the other elite subsystem or a non-elite subsystem. Columns 5 and 6 show results inconsistent with sibling competition: effects on first choice demand are actually 4.0 to 4.9 percentage points smaller when the school below the cutoff is not elite, and these differences are statistically significant. Columns 7 and 8 show negative and significant differential effects on the total number of schools chosen in the elite cutoff subsystem, providing additional evidence against the competition channel. There is a caveat to this evidence, however. Sibling pairs near an elite subsystem cutoff may be different than sibling pairs found near non-elite cutoffs. If the former pairs are on average much less competitive than the latter and thus react differently to older sibling admission, then the observed effect heterogeneity may be due to sample composition differences rather than differences in characteristics between the schools above and below the cutoff.

Third, according to the psychology literature reviewed in Joensen and Nielsen (2018), competition is more likely to manifest in same-sex pairings and when the siblings have similar levels of achievement. Table 6 repeats the above exercise separately for same- and opposite-sex subgroups and splitting the sample based on whether the absolute difference in middle school GPA is greater than the median of the absolute difference in the sample. Columns 1 through 4 and 9 through 12 show the results for subsamples where competition is expected to be most important, yet the pattern of results is similar to that in the full sample case. In particular, the point estimates on differential effects with respect to change in school cutoff is close to zero in each case, while the estimated differential effects with respect to the type of school below the cutoff continue to be inconsistent with those predicted by sibling competition.

Taken together, the evidence on differential responses to older sibling assignment with respect to school competitiveness strongly suggests that between-sibling competition and parental pressure are not driving the observed effects of older sibling assignment on school demand.

4.5 Channel: Path dependence in decision-making, salience, and information

This section explores the possibility that students follow the path of their older siblings due to incomplete information about schools and other difficulties in the decision-making process. For example, perhaps there is a high cognitive and time cost to investigating schools and ranking them, such that students (and their parents) find it easier to simply mimic the older sibling. Related to this, given that there are hundreds of possible schools from which to choose, perhaps the older sibling's school is made salient by his attendance there and the student is thus more likely to have it in mind when ranking schools. Finally, perhaps students are risk-averse, so that they prefer schools about which they have more information regarding school characteristics and student-school match quality. In this case, all else equal, students should prefer schools that older siblings have attended or that are observably similar to those schools, such as schools in the same subsystem.[21]

Table 7 presents evidence related to path dependence and salience effects by showing heterogeneous effects of older sibling admission on first choice demand. Older sibling admission is found to have a 2.7 p.p. higher effect when it is the firstborn child who is admitted, consistent with path dependence in which the oldest child sets a first example that is then followed by subsequent children. If this dependence is because students and parents are avoiding high decision-making costs, we might expect that effects are larger when parental education is lower, students are lower-achieving, or they come from relatively lower-quality middle schools. Column 2 finds a tightly-bounded zero differential effect with respect to parental education, proxied by an indicator for at least one parent having completed high school. Columns 3 shows a 1.6 p.p. *higher* effect for younger siblings with middle school GPAs above the median, while column 4 finds no evidence for a differential effect with respect to younger siblings' self-reported hours studied per week. Column 5 shows no evidence of a differential effect for younger siblings belonging to middle school cohorts with above-median COMIPEMS exam scores. Taken together, this evidence does not favor the hypothesis that students with lower socioeconomic backgrounds and lower achievement find it difficult to build a portfolio of school choices and rely more on mimicking older siblings. Furthermore, column 6 finds no evidence of a differential effect when the cutoff school is a "neighborhood school" within 2 kilometers of home.[22] This casts doubt on the importance of pure salience effects, to the extent that students are aware of the schools in their neighborhood.

The role of information about school characteristics and student-school match quality is difficult to test without observing actual information transmission between siblings. An imperfect proxy for learning is an indicator for whether the older sibling graduates from high school. The logic for using this proxy is as follows. One contributor to dropout is a bad

[21]The online appendix presents a formal model to make explicit the relationship between information and choice for risk-averse students.

[22]The finding is robust to alternative choices of the cutoff, and indeed, the marginal effect of admission is nearly invariant with respect to distance from home to school.

match between student and school. That is, there are students who will drop out from some schools but not others. Siblings are often similar in their preferences and abilities, so if the older sibling drops out, this suggests to the younger sibling that the school may not be a good match for him either.

Any estimates of differential admission effects with respect to dropout must be treated as suggestive rather than rigorously causal, because dropout is not randomly assigned (indeed, if it were, it would have no informational content for the student). Consider the reduced form heterogeneous effects specifications, found in equation 2, where x_i is now a dummy variable equal to 1 if the older sibling graduates. If dropout were randomly assigned, then $\widehat{\alpha}$ would give the additional average effect of admission when the older sibling graduates. The problem arises when $cor\left(x_i \times admit_{ijt}, \varepsilon_{ijt}\right) \neq 0$, so that students who are *differentially* more or less likely to drop out when admitted to the cutoff school are systematically more or less likely to be emulated, or have family characteristics that affect the likelihood of choosing the cutoff school.

The empirical analysis addresses the potential issue of endogenous heterogeneous effects in three ways. First, it considers multiple samples and outcomes and argues that the pattern in the results is consistent with learning. Second, it may be that the sibling admission effect is heterogeneous with respect to the school's graduation rate or other school characteristics, not the sibling's individual graduation outcome. The specification that allows cutoff-specific admission coefficients accounts for this, so that the admission-graduation interaction term gives the estimated heterogeneity due to sibling dropout *conditional on cutoff school characteristics*. Finally, the specification with cutoff-specific effects also controls for the older sibling's middle school grade point average, which is a significant predictor of high school dropout, and its interactions with admission and exam score.

Keeping in mind the caveats associated with using graduation status to proxy for a surprise to match quality, as well as the data limitations in using the graduation data, Table

8 shows that the admission effect is heterogeneous with respect to older sibling dropout.[23] The effect of admission on same-school demand is higher when the older sibling graduates. Column 1 gives the differential effect of admission on application to the cutoff school with respect to graduation status. On average, admission has a 2.4 p.p. higher impact on first choice preference for the cutoff school when the older sibling graduates.[24] The differential effect is illustrated in Figure 6. Column 2 controls for the interaction between admission and older sibling GPA while estimating separate admission coefficients for each school-by-older sibling exam year, with similar results.

In order to further explore the issue of endogenous differential dropout, columns 3 and 4 present estimates of the impact on the first non-elite choice of students whose siblings were near the cutoff of a non-elite school. This, in part, addresses the possibility that students whose older siblings are more able to graduate in the cutoff school are more likely to choose better schools. In particular, one might worry that older siblings able to graduate from elite schools are from families with high academic expectations who push the younger sibling to apply as well. Focusing on the non-elite preferences of students with siblings at non-elite cutoffs, we are likely to mitigate this confounding factor to some degree. The differential effect here is large, 7.5 p.p. compared to a rate of 17% for students scoring one point below the cutoff. Adding controls in column 4, the estimates are almost identical.

There is also heterogeneity in the effect on demand for other schools in the same subsystem. Columns 5 and 6 report the differential impact on the number of other schools selected in the cutoff school's subsystem, restricting the sample to cases where the older sibling is at the margin of a subsystem. The point estimates suggest that older sibling graduation

[23]Sample size is smaller than in the previous analysis, due to the fact that graduation data only exist for older siblings from the 2005 to 2008 cohorts and that graduation outcomes are missing for students at the UNAM schools, which are highly-demanded as first choices. This necessitates inclusion of all sibling pairs 1 to 5 years apart in age in the estimation sample. A sibling one year below his older sibling still has most of an academic year to learn about his sibling's school, since school begins in the early fall and preference listings are not due until February or March. Although graduation has not occurred yet for the siblings who are 1 or 2 years apart, in Mexico City most dropout occurs in the first year.

[24]The estimates imply a smaller average effect of admission than did previous tables. This is because the UNAM cutoff schools are missing from the sample, and much of the admission effect on first choice demand comes from the elite UNAM and IPN subsystems.

increases the number of other schools chosen in the same subsystem by 0.12. The same-school and subsystem effects are consistent with students reacting to older siblings' "good" and "bad" outcomes differently, learning about match quality and updating their choice behavior accordingly.

4.6 Effect on school assignment outcomes

Older sibling admission outcomes affect not only the stated preferences of younger siblings, but also their realized school assignments. Table 9, accompanied by Figure 7, shows the admission effect on the younger sibling's assignment. Admission to the cutoff school increases the probability of the younger sibling's assignment to the same school by 4.5 p.p., a large increase over the 7.1% assignment rate for students one point below the admission cutoff. A similar result holds for the school immediately below the cutoff, with an admission effect of 5.1 p.p. Columns 3 and 4 limit the sample to observations where the older sibling has schools in different subsystems above and below the cutoff. Older sibling admission increases the probability of being assigned to any school within the cutoff subsystem by 4.4 p.p., similar to the 4.6 p.p. effect of assignment to the subsystem below the cutoff. Appendix Table B.1 shows that the average effect of older sibling admission to the cutoff school or subsystem has a tightly-bounded null effect on younger sibling COMIPEMS exam scores, indicating that these effects on assignment are driven by changes in reported preferences rather than higher assignment priority.

Perhaps most interesting is the effect that older sibling admission to elite high schools has on younger sibling preference for, and assignment to, elite schools. Among the students least likely to apply to elite schools, older sibling admission has a large effect on both elite application and assignment rates. To show this, I first use a probit model to estimate the determinants of elite first choice among younger siblings of students who applied to elite schools and were rejected. Probit estimates are in Appendix Table B.2. [25] These estimates

[25] The most important predictor of elite preference is the proportion of the older sibling's middle school cohort expressing elite preference. This measure captures the preferences, constraints, and peer network effects in the older sibling's cohort, which are relevant for most younger siblings as they often attend the

are then used to generate a predicted probability of elite first choice for all younger siblings of students who were either above or below the cutoff for elite admission. Table 10, Panel A shows estimates of the sibling admission effect on students falling within different bins of the predicted elite first choice probability. Column 1 estimates a large 12.7 p.p. increase in the probability of choosing an elite school among students with less than a 60% predicted probability of doing so. As the predicted probability increases, the marginal effect of older sibling admission falls dramatically. Panel B shows similar results for the total number of elite schools selected.

The consequent effect on elite *assignment* (Panel C) is 3.5 p.p. among those least likely to choose an elite school, compared to an estimated zero effect for those most likely to express elite preference. This represents a 19% increase in the probability of admission, compared to the rate of elite admission among those with older siblings missing elite assignment by one point (18%). This implies that there exist many students who are capable of elite admission but only apply when they are sufficiently exposed to an elite school through their siblings.[26]

4.7 Validity of the RD design

This section presents two standard checks for the validity of the RD design. The first is whether the density of the running variable (centered COMIPEMS score) suddenly increases or decreases when it crosses the cutoff, as suggested by McCrary (2008). This might occur if the younger siblings of rejected students were less likely to apply to high school, for example if rejected students were more likely to drop out of school and younger siblings followed that example. Another, less likely possibility is that admission induces behavior that makes it impossible to match siblings to each other, such as changing their phone number or middle school. Figure 8, Panel A shows the density of centered COMIPEMS score for the RD sample of older siblings. There is no clear change in density across the threshold. Panel B

same middle school or live in the same neighborhoods as these students.

[26]It was shown earlier that there is little evidence for the sibling competition and parental expectations channels. The differential effects on choice and assignment thus seem to be driven by the more generic path dependence and information channels, in the sense that they apply to non-elite schools and subsystems as well.

gives a closer view of the density near the cutoff. Implementing McCrary's formal test for a discontinuity in the density yields an estimated difference in log height of 0.006 ($SE = 0.007$) at the cutoff, failing to reject continuity.

The second check is to repeat the reduced form RD regressions, this time using exogenous student characteristics as the dependent variables. Imbens and Lemieux (2008) propose this as a way of verifying that exogenous characteristics do not suddenly change at the cutoff, which would call into question whether the endogenous variable would be balanced in the absence of a treatment effect. In order to jointly test that the admission coefficients are zero for all tested exogenous characteristics, seemingly unrelated regression (Zellner 1962) is used. Table 11 shows the results, failing to reject that the admission coefficients are equal to zero ($p = 0.86$). The point estimates are quite precise as well, ruling out even fairly small covariate imbalances. The absence of visible discontinuities in covariates is illustrated in Figure 9. Thus both checks yield support for the validity of the RD design.

5 Discrete choice model of school choice

In this section, the basic RD design is extended to a discrete choice model of school choice. It allows for a natural parameterization of the impact of sibling admission: the change in willingness to travel to that school or another school in the same subsystem, which with further assumptions can then be translated into a willingness to pay measure. It is also used to address a potential causal channel for the admission effect where preference shifts toward schools *close* to the older sibling's school, for example if the family sets a commuting pattern that includes that school.

5.1 Method

The RD design is incorporated into a discrete choice model by estimating the following equation, expressing expected utilities from each school as a function of school characteristics and older sibling assignment (suppressing time subscripts):

$$\begin{aligned} U_{ij}^{*} = \theta cut_{ij} + \delta \left(cut_{ij} \times admit_{i}\right) + \beta_{1} \left(\tilde{s}_{i} \times cut_{ij}\right) + \beta_{2} \left(\tilde{s}_{i} \times cut_{ij} \times admit_{i}\right) + \\ \underline{\theta} below_{ij} + \underline{\delta} \left(below_{ij} \times admit_{i}\right) + \underline{\beta}_{1} \left(\tilde{s}_{i} \times below_{ij}\right) + \underline{\beta}_{2} \left(\tilde{s}_{i} \times below_{ij} \times admit_{i}\right) + \\ \gamma dist_{ij} + \varepsilon_{ij} \end{aligned} \tag{3}$$

where $cut_{ij} = 1$ when student i's sibling is in school j's cutoff sample and 0 otherwise, $admit_i$ is a dummy for whether the older sibling meets the cutoff score, $below_{ij}$ is a dummy for whether the school is immediately below the student's cutoff school, and $dist_{ij}$ is the straight-line distance between the centroid of the student's postal code of residence and school j. Allowing expected utility to be higher or lower for cutoff schools and the school assigned for scores immediately below the cutoff (through θ and $\underline{\theta}$, respectively), and for this expected utility to vary around the cutoff, δ and $\underline{\delta}$ capture only the discontinuous jumps in expected utility caused by the older sibling crossing the cutoff score and being admitted to the school above the cutoff instead of the school immediately below.

Incorporating subsystems into the model is straightforward. For the M different subsystems, let $X_j^1, ..., X_j^M$ be dummy variables equal to 1 if school j belongs to the corresponding subsystem and 0 otherwise. Define $cutsub_{ij}$ equal to 1 if j belongs to the cutoff school's subsystem and $belowsub_{ij}$ equal to 1 if j belongs to the "below" school's subsystem, 0 otherwise. Adding these variables into the RD specification, we have:

$$\begin{aligned} U_{ij}^{*} = \theta cut_{ij} + \delta \left(cut_{ij} \times admit_{i}\right) + \beta_{1} \left(\tilde{s}_{i} \times cut_{ij}\right) + \beta_{2} \left(\tilde{s}_{i} \times cut_{ij} \times admit_{i}\right) + \\ \underline{\theta} below_{ij} + \underline{\delta} \left(below_{ij} \times admit_{i}\right) + \underline{\beta}_{1} \left(\tilde{s}_{i} \times below_{ij}\right) + \underline{\beta}_{2} \left(\tilde{s}_{i} \times below_{ij} \times admit_{i}\right) + \\ \sum_{\ell=2}^{M} X_{j}^{\ell} \left(\pi^{\ell} + \eta^{\ell} cutsub_{ij} + \underline{\eta}^{\ell} belowsub_{ij}\right) + \\ cutsub_{ij} \left[\phi admit_{i} + \psi_{1} \tilde{s}_{i} + \psi_{2} \left(\tilde{s}_{i} \times admit_{i}\right)\right] + \\ belowsub_{ij} \left[\underline{\phi} admit_{i} + \underline{\psi}_{1} \tilde{s}_{i} + \underline{\psi}_{2} \left(\tilde{s}_{i} \times admit_{i}\right)\right] + \gamma dist_{ij} + \varepsilon_{ij}. \end{aligned} \tag{4}$$

This specification allows marginal expected utilities to vary depending on whether the cutoff school belongs to j's subsystem, the older sibling's centered exam score, and whether the sibling exceeded the cutoff score (ϕ, the coefficient of interest). The corresponding underlined coefficients are all analogous except that they apply to the subsystem of the school attended by the older sibling if he scores below the cutoff.

Assuming that ε_{ij} is distributed i.i.d. extreme value type I, the parameters of this model can be estimated with a conditional logit, where the outcome variable is selecting the school as the first choice.[27] The bandwidth is set to 10 points on either side of the cutoff. The bandwidth choice here is arbitrary, but is within the range of bandwidths chosen by the IK procedure in the reduced form analysis. As in the reduced form analysis, the edge kernel is used to weight observations with respect to centered COMIPEMS score. Results from an exploded logit model, modeling the joint choice of the student's first three options, are also presented.[28]

In order to improve the fit of the model, controls are included for mean COMIPEMS score of students assigned to the school in the previous year and proportion of the older sibling's middle school cohort choosing the school as first choice. These controls are intended to capture a wide variety of unobserved factors that predict demand for the school and the corresponding preference parameters are not given a structural interpretation. Additionally, some specifications include controls for distance of the school from the schools above and below the cutoff, as well as the interactions of these measures with older sibling admission to the cutoff school and piecewise-linear terms in centered COMIPEMS score.

5.2 Results

The conditional and exploded logit results have signs consistent with the reduced form estimates. Appendix Table B.3 provides estimated parameters from the logit specification in equation 4. Table 12 presents willingness to travel (WTT) estimates for the sibling

[27]As in the reduced form analysis, the student is considered to have chosen a particular school if he selects any career track within it.

[28]Hastings, Kane, and Staiger (2009) estimate an exploded mixed logit model of school choice. The model used here does not have random coefficients but is otherwise similar.

admission effects, calculated by taking the negative of the ratio of the admission coefficient to the distance from home coefficient.[29]

Older sibling admission to the cutoff school increases expected utility from that school and reduces the expected utility from the school below the cutoff. The conditional logit results imply an increase in WTT of 3.5 km (one-way) for the school above the cutoff and a 5.8 km decrease for the school below the cutoff. Exploded logit results, modeling the joint probability of the first three schools chosen, find slightly lower WTT effects of admission, implying changes of 3.0 km for the school above the cutoff and 3.3 for the school below. When the older sibling is on the margin between one *subsystem* and another (columns 3 and 4), admission to the school above the cutoff increases WTT to all other schools in that subsystem by between 1.2 and 1.4 km. WTT for other schools in the subsystem of the school below the cutoff fall by 0.9 to 1.0 km when the older sibling scores too high to be admitted there.[30] Columns 5 and 6 show that these effects, both on same-school and same-subsystem demand, are essentially unchanged when allowing for admission to affect demand for schools on the basis of how far they are located from the cutoff school. Thus it seems that admission has effects that go far beyond proximity and commuting considerations.

Additional assumptions allow for interpretation of the effect sizes as measures of marginal willingness to pay (WTP). Taking the average WTT effect between the schools above and below the cutoff in the full sample exploded logit specification, we have a 3.2 km average increase in WTT due to sibling admission. But students must travel both to and from school, so this measure should be doubled to 6.4 km/day. Students in Mexico have 195 instructional days per year, so the annualized effect on WTT is $6.4 * 195 = 1,248$ km/year. Translating this measure to travel time is difficult because students travel using a combination of subway, private bus, driving, and walking. Assuming that the average speed of travel over these modes during rush hour in Mexico City is 10 km/hour, then students are willing to spend $1,248/10 =$

[29]Standard errors for WTT effects are computed using the delta method.

[30]The total change in WTT for the cutoff school when the student is at the boundary of a subsystem is obtained by summing the effect of admission to the school with the effect for admission to the subsystem.

124.8 additional hours per year traveling as a result of sibling admission. According to the 2011 National Survey of Occupation and Employment (Instituto Nacional de Estadística y Geografía 2011), the average urban teen wage was $2 USD/hour. Taking this as the average valuation of time for students in the estimation sample, the change in WTP due to sibling admission is $249.60/year. High school is three years long in Mexico City, so the total effect on WTP for a student completing school is $749. This is likely to be a conservative estimate because traveling farther may require paying an additional bus fare of about $.50/day. The $749 estimate is equal to about two months of self-reported family income in this sample, although such self-reports by adolescents may be unreliable. As a point of comparison, $749 was approximately the median monthly household income among urban Mexican households in 2010 (Instituto Nacional de Estadística y Geografía 2010). Thus the additional value placed on a school due to sibling admission appears quite large for the typical family facing this decision.

6 Conclusion

Older siblings' school assignments strongly affect students' stated preferences and assignment outcomes, an effect that is not driven by convenience or costs of attendance. The findings here provide support for a causal interpretation of the correlation observed in Goodman et al. (2015). They also bolster the assertion in Hoxby and Avery (2012) that older peers matter for school choice.

One implication of the results is that any policy affecting student school assignments will also spill over onto younger siblings' school preferences. This implies, for example, a potential delayed multiplier effect of affirmative action systems in school admissions. But other policies may also result in spillovers, such as information campaigns in public school systems or marketing and recruiting efforts by private schools. The fact that siblings affect preferences not only for their exact same school, but also for similar schools, suggests that outreach and marketing efforts by specific public or private schools may be affecting demand for their respective sectors in addition to their own schools.

Given that the observed effects do not appear to act through channels that are specific to siblings—convenience, commuting, or sibling rivalry—the findings raise two avenues for further investigation. First, understanding the role of broader peer networks, including older friends and classmates, in determining school choice behavior is a natural next step. Second, the extent to which information-sharing and path dependence due to the perceived difficulty of navigating the school choice process are the key causal channels in affecting choice needs to be better understood. If students rely on their networks because they lack information about school characteristics or student-school fit, this suggests a role for greater levels of information provision by public school systems and other organizations. Furnishing printed material containing data beyond school-level aggregates is one potential way to begin, as in Hoxby and Turner (2013) in the context of college choice where information on net costs is the predominant barrier. It is less clear what the informational constraints are in Mexico City and elsewhere, and whether these relate more to school-level characteristics or how well a school would suit a particular student. In the latter case, family and wider peer networks may be indispensable in helping students and their parents to choose the right school for them.

References

Aaronson, Daniel, Lisa Barrow, and William Sander. 2007. "Teachers and Student Achievement in the Chicago Public High Schools." *Journal of Labor Economics* 25 (1): 95–135.

Abdulkadiroglu, Atila, Joshua D. Angrist, and Parag A. Pathak. 2014. "The Elite Illusion: Achievement Effects at Boston and New York Exam Schools." *Econometrica* 82 (1): 137–196.

Abdulkadiroglu, Atila, and Tayfun Sonmez. 2010. "Matching Markets: Theory and Practice." Prepared for the Econometric Society World Congress, China.

Ajayi, Kehinde F. 2013. "School Choice and Educational Mobility: Lessons from Secondary School Applications in Ghana." Unpublished working paper, Boston University.

Angrist, Joshua, et al. 2002. "Vouchers for Private Schooling in Colombia: Evidence from a Randomized Natural Experiment." *American Economic Review* 92 (5): 1535–1558.

Ball, Stephen J., and Carol Vincent. 1998. "'I Heard It on the Grapevine': 'Hot' Knowledge and School Choice." *British Journal of Sociology of Education* 19 (3): 377–400.

Burgess, Simon, et al. 2014. "What Parents Want: School Preferences and School Choice." *Economic Journal* 125 (587): 1262–1289.

Cameron, A. Colin, Jonah B. Gelbach, and Douglas L. Miller. 2008. "Bootstrap-Based Improvements for Inference with Clustered Errors." *Review of Economics and Statistics* 90 (3): 414–427.

Cameron, A. Colin, and Douglas L. Miller. 2015. "A Practitioner's Guide to Cluster-Robust Inference." *Journal of Human Resources* 50 (2): 317–372.

Ceja, Miguel. 2006. "Understanding the Role of Parents and Siblings as Information Sources in the College Choice Process of Chicana Students." *Journal of College Student Development* 47 (1): 87–104.

Chade, Hector, and Lones Smith. 2006. "Simultaneous Search." *Econometrica* 74 (5): 1293–1307.

Chumacero, Romulo A., Daniel Gomez, and Ricardo D. Paredes. 2011. "I Would Walk 500 Miles (If It Paid): Vouchers and School Choice in Chile." *Economics of Education Review* 30 (5): 1103–1114.

Clark, Damon. 2010. "Selective Schools and Academic Achievement." *The B.E. Journal of Economic Analysis & Policy (Advances)* 10 (1): Article 9.

Dobbie, Will, and Roland G. Fryer Jr. 2011. "Are High-Quality Schools Enough to Increase Achievement Among the Poor? Evidence from the Harlem Children's Zone." *American Economic Journal: Applied Economics* 3 (3): 158–187.

Dubins, L. E., and D. A. Freedman. 1981. "Machiavelli and the Gale-Shapley Algorithm." *The American Mathematical Monthly* 88 (7): 485–495.

Dustan, Andrew, Alain de Janvry, and Elisabeth Sadoulet. 2017. "Flourish or Fail? The Risky Reward of Elite High School Admission in Mexico City." *Journal of Human Resources* 52 (3): 756–799.

Elias McAllister, Dora. 2012. "Mexican American First-Generation Students Perceptions of Siblings and Additional Factors Influencing their College Choice Process." PhD thesis, University of Maryland.

Estrada, Ricardo, and Jeremie Gignoux. 2017. "Benefits to Elite Schools and the Formation of Expected Returns to Education: Evidence from Mexico City." *European Economic Review* 95:168–194.

Friesen, Jane, et al. 2012. "How Do School 'Report Cards' Affect School Choice Decisions?" *Canadian Journal of Economics* 45 (2): 784–807.

Goodman, Joshua, et al. 2015. "The Relationship Between Siblings' College Choices: Evidence from One Million SAT-Taking Families." *Economics of Education Review* 48:75–85.

Hastings, Justine, and Jeffrey Weinstein. 2008. "Information, School Choice, and Academic Achievement: Evidence from Two Experiments." *Quarterly Journal of Economics* 123 (4): 1373–1414.

Hastings, Justine S., Thomas J. Kane, and Douglas O. Staiger. 2009. "Heterogeneous Preferences and the Efficacy of Public School Choice." NBER Working Paper No. 12145.

Hoop, Jacobus de. 2011. "Selective Schools and Education Decisions: Evidence from Malawi." Unpublished working paper, Free University of Amsterdam.

Hoxby, Caroline M., and Christopher Avery. 2012. "The Missing 'One-Offs': The Hidden Supply of High-Achieving, Low Income Students." NBER Working Paper No. 18586.

Hoxby, Caroline M., and Sarah Turner. 2013. "Expanding College Opportunities for High-Achieving, Low Income Students." SIEPR Discussion Paper No. 12-014.

Hsieh, Chang-Tai, and Miguel Urquiola. 2006. "The Effects of Generalized School Choice on Achievement and Stratification: Evidence from Chile's School Voucher Program." *Journal of Public Economics* 90 (8-9): 1477–1503.

Imbens, Guido, and Karthik Kalyanaraman. 2012. "Optimal Bandwidth Choice for the Regression Discontinuity Estimator." *Review of Economic Studies* 79 (3): 933–959.

Imbens, Guido W., and Thomas Lemieux. 2008. "Regression Discontinuity Designs: A Guide to Practice." *Journal of Econometrics* 142 (2): 615–635.

Instituto Nacional de Estadística y Geografía. 2010. *Encuesta Nacional de Ingresos y Gastos de los Hogares*. Available from http://www.beta.inegi.org.mx/proyectos/enchogares/regulares/enigh/nc/2010/.

— . 2011. *Encuesta Nacional de Ocupación y Empleo*. Available from http://www.beta.inegi.org.mx/proyectos/enchogares/historicas/enoe/.

Jackson, Kirabo. 2010. "Do Students Benefit From Attending Better Schools?: Evidence From Rule-based Student Assignments in Trinidad and Tobago." *The Economic Journal* 142 (549): 1399–1429.

Joensen, Juanna Schroter, and Helena Skyt Nielsen. 2018. "Spillovers in Education Choice." Forthcoming, *Journal of Public Economics*.

Kaczynski, Kristen M. 2011. "Exploring the Influence of Siblings and their Relationships on the College Choice Process." PhD thesis, George Washington University.

Koning, Pierre, and Karen van de Wiel. 2013. "Ranking the Schools: How Quality Information Affects School Choice in the Netherlands." *Journal of the European Economic Association* 11 (2): 466–493.

Lai, Fang, Elisabeth Sadoulet, and Alain de Janvry. 2009. "The Adverse Effects of Parents' School Selection Errors on Academic Achievement: Evidence from the Beijing Open Enrollment Program." *Economics of Education Review* 28 (4): 485–496.

— . 2011. "The Contributions of School Quality and Teacher Qualifications to Student Performance: Evidence from a Natural Experiment in Beijing Middle Schools." *Journal of Human Resources* 46 (1): 123–53.

Lee, David S., and David Card. 2008. "Regression Discontinuity Inference with Specification Error." *Journal of Econometrics* 142 (2): 655–674.

Lucas, Adrienne M., and Isaac M. Mbiti. 2012. "The Determinants and Consequences of School Choice Errors in Kenya." *American Economic Review: Papers and Proceedings* 102 (3): 283–288.

— . 2014. "Effects of School Quality on Student Achievement: Discontinuity Evidence from Kenya." *American Economic Journal: Applied Economics* 6 (3): 234–63.

Manski, Charles F. 1993. "Identification of Endogenous Social Effects: The Refection Problem." *Review of Economic Studies* 60 (3): 531–42.

— . 1995. *Identification Problems in the Social Sciences*. Harvard University Press.

McCrary, Justin. 2008. "Manipulation of the Running Variable in the Regression Discontinuity Design: A Density Test." *Journal of Econometrics* 142 (2): 698–714.

Mizala, Alejandra, and Miguel Urquiola. 2013. "School Markets: The Impact of Information Approximating Schools' Effectiveness." *Journal of Development Economics* 103:313–335.

Muralidharan, Karthik, and Venkatesh Sundararaman. 2015. "The Aggregate Effect of School Choice: Evidence from a Two-Stage Experiment in India." *Quarterly Journal of Economics* 130 (3): 1011–66.

Pop-Eleches, Cristian, and Miguel Urquiola. 2013. "Going to a Better School: Effects and Behavioral Responses." *American Economic Review* 103 (4): 1289–1324.

Roth, Alvin E. 1982. "Incentive Compatibility in a Market With Indivisible Goods." *Economics Letters* 9 (2): 127–132.

Zellner, Arnold. 1962. "An Efficient Method of Estimating Seemingly Unrelated Regressions and Tests for Aggregation Bias." *Journal of the American Statistical Association* 57 (298): 348–368.

Zhang, Hongliang. 2016. "Identification of Local Average Treatment Effects under Imperfect Matching with an Application to Chinese Elite Schools." *Journal of Public Economics* 142:56–82.

Tables

Table 1: Summary statistics for full and sibling samples

Panel A. Student-level summary statistics	(1) All students	(2) Matched younger sibling sample	(3) Younger sib observations in RD sample (BW = 10)
Male	0.49 (0.50)	0.48 (0.50)	0.48 (0.50)
Maximum of mother's and father's education (years)	10.39 (3.56)	10.35 (3.39)	10.61 (3.36)
Number of siblings	2.18 (1.45)	2.45 (1.34)	2.37 (1.29)
Birth order (1 is firstborn)	2.17 (1.35)	2.73 (1.08)	2.68 (1.05)
Hours studied per week	4.86 (3.18)	4.89 (3.17)	5.00 (3.19)
Middle school grade point average (of 10)	8.06 (0.82)	8.14 (0.82)	8.15 (0.82)
Number of schools ranked	9.32 (3.90)	8.94 (3.55)	9.37 (3.63)
Elite school as first choice	0.63 (0.48)	0.64 (0.48)	0.72 (0.45)
Assigned to an elite school	0.21 (0.41)	0.20 (0.40)	0.23 (0.42)
COMIPEMS examination score	63.05 (19.34)	61.81 (18.83)	62.56 (18.44)
Distance from student's home postal code to first choice school (km)	7.68 (6.22)	7.87 (6.18)	8.13 (6.12)
Distance from student's home postal code to first non-elite choice school (km)	6.12 (5.36)	5.96 (5.05)	6.05 (5.04)
Took HS ENLACE (2005-2008 cohort students assigned to non-UNAM schools)	0.54 (0.50)	0.59 (0.49)	0.59 (0.49)
Observations	3051216	540118	480693

Panel B. Sibling pair-level summary statistics	(1) Matched younger sibling sample	(2) Younger sib observations in RD sample (BW = 10)
Grade year difference between siblings	2.76 (1.47)	2.76 (1.47)
Same sex sibling pair	0.51 (0.50)	0.51 (0.50)
Younger sibling MS GPA - older sibling MS GPA	−0.03 (0.93)	−0.07 (0.92)
Absolute value of MS GPA difference	0.73 (0.57)	0.73 (0.56)
Siblings have same first choice school	0.33 (0.47)	0.35 (0.48)
Observations	540118	480693

Note: Standard deviations in parentheses. Statistics in column 2 of Panel A are for younger siblings in the matched sample, described in the text. Each observation in the RD sample is a sibling pair-cutoff school pairing, such that a sibling pair appears multiple times in the sample if his older sibling is within the 10 point bandwidth of two different cutoff schools. The RD sample consists of 288,863 unique students appearing on average 1.66 times each.

Table 2: Effect of older sibling admission on younger sibling choice

	(1) Cutoff school is first choice	(2) Cutoff school is any choice	(3) School below cutoff is first choice	(4) School below cutoff is any choice
Score $\geq$ cutoff	0.073***	0.105***	−0.052***	−0.127***
	(0.0023)	(0.0028)	(0.0019)	(0.0038)
Observations	437995	480693	302643	302643
Adjusted R^2	0.158	0.144	0.031	0.113
Mean of DV 1 pt below cutoff	0.142	0.606	0.077	0.618
Bandwidth	9.5	10.3	7.0	6.9

Note: Dependent variables are dummy variables pertaining to the choice of the younger sibling. Regressions include cutoff school-year fixed effects and piecewise-linear polynomial terms in older sibling's centered exam score. Observations are weighted using the edge kernel. Standard errors accounting for clustering at the older sibling level are in parentheses. Stars for statistical significance are based on t-tests using (G-1) degrees of freedom, where G is the number of points of support of the centered score. * p<0.10, ** p<0.05, *** p<0.01.

Table 3: Effect of older sibling admission on younger sibling's preference for same school, heterogeneity by grade year difference of siblings

	(1) Cutoff school is first choice	(2) Cutoff school is first choice	(3) Cutoff school is any choice	(4) Cutoff school is any choice	(5) School below cutoff is first choice	(6) School below cutoff is first choice	(7) School below cutoff is any choice	(8) School below cutoff is any choice
Score $\geq$ cutoff	0.078***	(One per	0.106***	(One per	−0.060***	(One per	−0.125***	(One per
	(0.0033)	cutoff)	(0.0039)	cutoff)	(0.0028)	cutoff)	(0.0052)	cutoff)
(Score $\geq$ cutoff) $\times$ (Sibs 3+ years apart)	−0.009*	−0.011**	−0.003	−0.004	0.016***	0.017***	−0.003	−0.004
	(0.0047)	(0.0047)	(0.0056)	(0.0056)	(0.0038)	(0.0038)	(0.0075)	(0.0075)
Observations	437995	437981	480693	480681	302643	302617	302643	302617
Adjusted R^2	0.159	0.162	0.151	0.153	0.032	0.037	0.118	0.119
Mean of DV 1 pt below cutoff	0.142	0.142	0.606	0.606	0.077	0.077	0.618	0.618
Bandwidth	9.5	9.5	10.3	10.3	7.0	7.0	6.9	6.9

Note: Regressions include cutoff school-year fixed effects, piecewise-linear polynomial terms in older sibling's centered exam score, a dummy variable for siblings 3+ years apart in exam years, and interactions between this dummy and the piecewise-linear polynomial. Observations are weighted using the edge kernel. Standard errors accounting for clustering at the older sibling level are in parentheses. Stars for statistical significance are based on t-tests using (G-1) degrees of freedom, where G is the number of points of support of the centered score. * p<0.10, ** p<0.05, *** p<0.01.

Table 4: Effect of older sibling admission on number of other schools chosen in cutoff subsystem

	(1) Schools chosen in subsystem of cutoff school	(2) Schools chosen in subsystem of cutoff school	(3) Schools chosen in subsystem of cutoff school	(4) Schools chosen in subsystem of school below cutoff	(5) Schools chosen in subsystem of school below cutoff	(6) Schools chosen in subsystem of school below cutoff
Score ≥ cutoff	0.206***	0.198***	(One per	−0.158***	−0.158***	(One per
	(0.0150)	(0.0204)	cutoff)	(0.0126)	(0.0178)	cutoff)
(Score ≥ cutoff) × (Sibs 3+ years apart)		0.015	0.017		−0.006	−0.004
		(0.0300)	(0.0301)		(0.0250)	(0.0252)
Observations	305440	305440	305416	362747	362747	362733
Adjusted R^2	0.206	0.211	0.214	0.055	0.059	0.060
Mean of DV 1 pt below cutoff	1.867	1.867	1.867	1.421	1.421	1.421
Bandwidth	13.4	13.4	13.4	16.3	16.3	16.3

Note: Dependent variables are the number of schools in the respective subsystems chosen by the younger sibling, excluding either the cutoff school or the school immediately below the cutoff. Samples are limited to older siblings for whom rejection from the cutoff school results in admission to a school in a different subsystem. Regressions include cutoff school-year fixed effects and piecewise-linear polynomial terms in older sibling's centered exam score. Specifications in columns 2 and 4 also include a dummy variable for siblings 3+ years apart in exam years and interactions between this dummy and the piecewise-linear polynomial. Observations are weighted using the edge kernel. Standard errors accounting for clustering at the older sibling level are in parentheses. Stars for statistical significance are based on t-tests using (G-1) degrees of freedom, where G is the number of points of support of the centered score. * p<0.10, ** p<0.05, *** p<0.01.

Table 5: Effect of older sibling admission on younger sibling's choices, heterogeneity by differences in schools above and below cutoff

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Sample	Cutoff school and school below cutoff in same subsystem				Cutoff school elite; school below cutoff in different subsystem			
Dependent variable	Cutoff school is first choice	Cutoff school is first choice	Cutoff school is any choice	Cutoff school is any choice	First choice belongs to cutoff school subsystem	First choice belongs to cutoff school subsystem	Schools chosen in cutoff school subsystem	Schools chosen in cutoff school subsystem
Score ≥ cutoff	0.077***	(One per	0.075***	(One per	0.153***	(One per	0.424***	(One per
	(0.0045)	cutoff)	(0.0054)	cutoff)	(0.0115)	cutoff)	(0.0484)	cutoff)
(Score ≥ cutoff) × (Dif. in cutoff scores of schools above and below cutoff ≥ median)	0.001	−0.004	−0.005	0.011				
	(0.0064)	(0.0068)	(0.0076)	(0.0083)				
(Score ≥ cutoff) × (School below cutoff belongs to non-elite subsystem)					−0.049***	−0.040***	−0.125**	−0.141**
					(0.0128)	(0.0130)	(0.0543)	(0.0557)
Observations	238017	237977	258930	258892	144313	144313	175162	175162
Adjusted R^2	0.200	0.204	0.173	0.175	0.168	0.169	0.120	0.121
Mean of DV 1 pt below cutoff	0.163	0.163	0.643	0.643	0.563	0.563	2.284	2.284
Bandwidth	10.7	10.7	11.3	11.3	12.4	12.4	15.8	15.8

Note: Regressions include cutoff school-year fixed effects, piecewise-linear polynomial terms in older sibling's centered exam score, a dummy variable for the stated interaction variable (difference in cutoff scores of schools above and below cutoff > median or school below cutoff belongs to non-elite subsystem), and interactions between this dummy and the piecewise-linear polynomial and fixed effects. Observations are weighted using the edge kernel. Standard errors accounting for clustering at the older sibling level are in parentheses. Stars for statistical significance are based on t-tests using (G-1) degrees of freedom, where G is the number of points of support of the centered score. * p<0.10, ** p<0.05, *** p<0.01.

Table 6: Separating effects by sibling pair characteristics, heterogeneity by sibling pair characteristics and differences in schools above and below cutoff

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Sample	Same-sex sibling pair				Opposite-sex sibling pair			
Subsample	Cutoff school and school below cutoff in same subsystem		Cutoff school elite; school below cutoff in different subsystem		Cutoff school and school below cutoff in same subsystem		Cutoff school elite; school below cutoff in different subsystem	
Dependent variable	Cutoff school is first choice	Cutoff school is first choice	Cutoff school is any choice	Cutoff school is any choice	First choice belongs to cutoff school subsystem	First choice belongs to cutoff school subsystem	Schools chosen in cutoff school subsystem	Schools chosen in cutoff school subsystem
Score $\geq$ cutoff	(One per cutoff)	(One per cutoff)	(One per cutoff)	(One per cutoff)	(One per cutoff)	(One per cutoff)	(One per cutoff)	(One per cutoff)
(Score $\geq$ cutoff) $\times$ (Dif. in cutoff scores of schools above and below cutoff $\geq$ median)	−0.004	0.006			−0.006	0.010		
	(0.0099)	(0.0104)			(0.0089)	(0.0113)		
(Score $\geq$ cutoff) $\times$ (School below cutoff belongs to non-elite subsystem)			−0.041***	−0.113*			−0.025	−0.201**
			(0.0112)	(0.0594)			(0.0199)	(0.0861)
Observations	132379	153007	160620	139532	145488	135691	59710	75481
Adjusted R^2	0.205	0.170	0.164	0.125	0.198	0.192	0.171	0.122
Mean of DV 1 pt below cutoff	0.172	0.664	0.571	2.291	0.152	0.620	0.554	2.275
Bandwidth	11.0	13.7	36.6	27.8	13.1	12.6	10.8	13.7
	(9)	(10)	(11)	(12)	(13)	(14)	(15)	(16)
Sample	Absolute difference in siblings' middle school GPA < median				Absolute difference in siblings' middle school GPA $\geq$ median			
Subsample	Cutoff school and school below cutoff in same subsystem		Cutoff school elite; school below cutoff in different subsystem		Cutoff school and school below cutoff in same subsystem		Cutoff school elite; school below cutoff in different subsystem	
Dependent variable	Cutoff school is first choice	Cutoff school is first choice	Cutoff school is any choice	Cutoff school is any choice	First choice belongs to cutoff school subsystem	First choice belongs to cutoff school subsystem	Schools chosen in cutoff school subsystem	Schools chosen in cutoff school subsystem
Score $\geq$ cutoff	(One per cutoff)	(One per cutoff)	(One per cutoff)	(One per cutoff)	(One per cutoff)	(One per cutoff)	(One per cutoff)	(One per cutoff)
(Score $\geq$ cutoff) $\times$ (Dif. in cutoff scores of schools above and below cutoff $\geq$ median)	0.009	0.009			−0.006	0.004		
	(0.0109)	(0.0122)			(0.0084)	(0.0098)		
(Score $\geq$ cutoff) $\times$ (School below cutoff belongs to non-elite subsystem)			−0.053**	−0.279***			−0.022	−0.089
			(0.0197)	(0.0881)			(0.0170)	(0.0657)
Observations	101135	118767	62363	71250	159128	167822	82448	118704
Adjusted R^2	0.193	0.181	0.170	0.116	0.220	0.190	0.182	0.134
Mean of DV 1 pt below cutoff	0.170	0.645	0.554	2.330	0.163	0.655	0.588	2.356
Bandwidth	10.9	12.4	12.3	14.8	15.6	16.7	15.0	24.5

Note: Regressions include cutoff school-year fixed effects, piecewise-linear polynomial terms in older sibling's centered exam score, a dummy variable for the stated interaction variable (difference in cutoff scores of schools above and below cutoff > median or school below cutoff belongs to non-elite subsystem), and interactions between this dummy and the piecewise-linear polynomial and fixed effects. Observations are weighted using the edge kernel. Standard errors accounting for clustering at the older sibling level are in parentheses. Stars for statistical significance are based on t-tests using (G-1) degrees of freedom, where G is the number of points of support of the centered score. * p<0.10, ** p<0.05, *** p<0.01.

Table 7: Differential effect of older sibling admission on school choice, by student and sibling pair characteristics

	(1)	(2)	(3)	(4)	(5)	(6)
Interaction variable	Older sibling is firstborn	Parent has high school education or above	Younger sibling has middle school GPA > median	Younger sibling has hours studied > median	Younger sib MS average COMIPEMS score > median	Distance from home to cutoff school < 2km
Dependent variable	Cutoff school is first choice	Cutoff school is first choice	Cutoff school is first choice	Cutoff school is first choice	Cutoff school is first choice	Cutoff school is first choice
(Score $\geq$ cutoff)	(One per cutoff)	(One per cutoff)	(One per cutoff)	(One per cutoff)	(One per cutoff)	(One per cutoff)
(Score $\geq$ cutoff) $\times$ (Interaction)	0.027***	−0.002	0.016***	0.008	−0.006	0.010
	(0.0052)	(0.0051)	(0.0050)	(0.0050)	(0.0048)	(0.0082)
Observations	356591	376261	392882	376987	430138	413348
Adjusted R^2	0.159	0.161	0.170	0.160	0.164	0.170
Mean of DV 1 pt below cutoff	0.144	0.145	0.144	0.145	0.142	0.141
Bandwidth	9.8	10.0	9.4	9.9	9.4	9.9

Note: Regressions include cutoff school-year fixed effects, piecewise-linear polynomial terms in older sibling's centered exam score, a dummy variable for the stated interaction variable (difference in cutoff scores of schools above and below cutoff > median or school below cutoff belongs to non-elite subsystem), and interactions between this dummy and the piecewise-linear polynomial and fixed effects. Observations are weighted using the edge kernel. Standard errors accounting for clustering at the older sibling level are in parentheses. Stars for statistical significance are based on t-tests using (G-1) degrees of freedom, where G is the number of points of support of the centered score. * p<0.10, ** p<0.05, *** p<0.01.

Table 8: Differential effect of older sibling admission on school choice by graduation outcome

	(1)	(2)	(3)	(4)	(5)	(6)
Sample	Full		Older siblings at non-elite cutoff		Older siblings at margin of subsystem admission	
Dependent variable	Cutoff school is first choice	Cutoff school is first choice	Cutoff school is first non-elite choice	Cutoff school is first non-elite choice	Schools chosen in cutoff school subsystem	Schools chosen in cutoff school subsystem
Score $\geq$ cutoff	0.021***	(One per	0.041***	(One per	0.115***	(One per
	(0.0046)	cutoff)	(0.0078)	cutoff)	(0.0408)	cutoff)
(Score $\geq$ cutoff) $\times$ (Older sibling graduated)	0.024***	0.023***	0.075***	0.079***	0.116**	0.120**
	(0.0065)	(0.0067)	(0.0105)	(0.0109)	(0.0549)	(0.0565)
Observations	142138	142138	105821	105821	75464	75453
Adjusted R^2	0.082	0.085	0.111	0.114	0.223	0.227
Mean of DV 1 pt below cutoff	0.085	0.085	0.165	0.165	1.519	1.519
Bandwidth	13.2	13.2	12.2	12.2	14.1	14.1

Note: Samples exclude all students at an UNAM school cutoff or who would attend an UNAM school if rejected from the cutoff school, since the UNAM schools have no graduation data available. Regressions include cutoff school-year fixed effects, piecewise-linear polynomial terms in older sibling's centered exam score, a dummy variable for whether the older sibling graduated, interactions between this dummy and the piecewise-linear polynomial and fixed effects, in addition to older sibling's GPA, a piecewise-linear polynomial in the interaction between older sibling's GPA and centered test score, and the interaction between GPA and admission. Observations are weighted using the edge kernel. Standard errors accounting for clustering at the older sibling level are in parentheses. Stars for statistical significance are based on t-tests using (G-1) degrees of freedom, where G is the number of points of support of the centered score. * p<0.10, ** p<0.05, *** p<0.01.

Table 9: Effect of older sibling admission on younger sibling assignment outcomes

	(1)	(2)	(3)	(4)
Sample	Full		Older siblings at margin of subsystem admission	
Dependent variable	Assigned to cutoff school	Assigned to school below cutoff	Assigned to school in subsystem of cutoff school	Assigned to school in subsystem below cutoff
Score $\geq$ cutoff	0.045***	−0.051***	0.044***	−0.046***
	(0.0016)	(0.0021)	(0.0032)	(0.0030)
Observations	644136	394299	285087	325166
Adjusted R^2	0.027	0.017	0.033	0.016
Mean of DV 1 pt below cutoff	0.071	0.116	0.181	0.224
Bandwidth	14.2	8.3	12.5	14.9

Note: Dependent variables are dummy variables pertaining to the admission outcome of the younger sibling. Samples for subsystem admission outcomes are limited to older siblings for whom rejection from the cutoff school results in admission to a school in a different subsystem. Regressions include cutoff school-year fixed effects and polynomials in older sibling's centered exam score. Observations are weighted using the edge kernel. Standard errors accounting for clustering at the older sibling level are in parentheses. Stars for statistical significance are based on t-tests using (G-1) degrees of freedom, where G is the number of points of support of the centered score. * p<0.10, ** p<0.05, *** p<0.01.

Table 10: Effect of older sibling admission to an elite school on younger sibling elite school choice and assignment

Panel A.	(1)	(2)	(3)
Dependent variable	Elite school as first choice		
Baseline predicted probability of selecting elite school as first choice	[0, .6]	(.6, .8]	(.8, 1]
Score ≥ cutoff	0.127***	0.073***	0.027***
	(0.0173)	(0.0092)	(0.0052)
Observations	13712	35223	53780
Adjusted R^2	0.051	0.030	0.022
Mean of DV 1 pt below cutoff	0.588	0.760	0.902
Bandwidth	18.4	14.7	16.6

Panel B.	(1)	(2)	(3)
Dependent variable	Number of elite schools chosen		
Baseline predicted probability of selecting elite school as first choice	[0, .6]	(.6, .8]	(.8, 1]
Score ≥ cutoff	0.584***	0.379***	0.198***
	(0.0971)	(0.0632)	(0.0536)
Observations	12898	38086	58770
Adjusted R^2	0.059	0.055	0.070
Mean of DV 1 pt below cutoff	2.408	3.839	5.550
Bandwidth	13.9	13.9	16.5

Panel C.	(1)	(2)	(3)
Dependent variable	Elite assignment		
Baseline predicted probability of selecting elite school as first choice	[0, .6]	(.6, .8]	(.8, 1]
Score ≥ cutoff	0.035**	0.035***	0.007
	(0.0138)	(0.0075)	(0.0076)
Observations	16162	56685	67556
Adjusted R^2	0.040	0.044	0.031
Mean of DV 1 pt below cutoff	0.176	0.222	0.302
Bandwidth	17.0	21.5	19.3

Note: Sample is composed of older siblings near the cutoff of an elite school who will be assigned to a non-elite school if they score below the cutoff. Columns partition this sample by the younger sibling's predicted probability of selecting an elite school as first choice. Regressions include cutoff school-year fixed effects and piecewise-linear polynomial terms in older sibling's centered exam score. Observations are weighted using the edge kernel. Standard errors accounting for clustering at the older sibling level are in parentheses. Stars for statistical significance are based on t-tests using (G-1) degrees of freedom, where G is the number of points of support of the centered score. * p<0.10, ** p<0.05, *** p<0.01.

Table 11: Test for balance of older sibling covariates

	(1) Parental education (years)	(2) Male	(3) Hours studied per week	(4) Middle school GPA	(5) Number of siblings	(6) Birth order (1=oldest)	(7) Mean COMIPEMS score of MS peers	(8) Same-sex sibling pair
Score ≥ cutoff	0.000	−0.002	0.015	0.003	0.002	0.007	0.019	−0.001
	(0.0186)	(0.0024)	(0.0195)	(0.0043)	(0.0072)	(0.0060)	(0.0382)	(0.0025)
Observations	517166	754586	448589	523443	525425	524033	523391	718786
Adjusted R^2	0.108	0.076	0.075	0.187	0.053	0.029	0.298	0.003
Mean of DV 1 pt below cutoff	10.648	0.432	5.080	8.174	2.322	1.674	63.054	0.519
SD of DV 1 pt below cutoff	3.323	0.495	3.233	0.783	1.268	1.044	7.662	0.500
Bandwidth	14.8	17.2	12.9	11.2	14.8	14.7	11.6	16.1
p-value for joint significance of Score ≥ cutoff coefficients								0.86

Note: Dependent variable corresponds to the older sibling. Regressions include cutoff school-year fixed effects and piecewise-linear polynomial terms in older sibling's centered exam score. Observations are weighted using the edge kernel. Standard errors accounting for clustering at the older sibling level are in parentheses. Stars for statistical significance are based on t-tests using (G-1) degrees of freedom, where G is the number of points of support of the centered score. * p<0.10, ** p<0.05, *** p<0.01.

Table 12: Willingness to travel estimates of admission for school attributes from conditional and exploded logit RD models of school choice

	(1)	(2)	(3)	(4)	(5)	(6)
Sample	Full		Older siblings at margin of subsystem admission			
Model	Conditional logit	Exploded logit	Conditional logit	Exploded logit	Conditional logit	Exploded logit
School above cutoff	3.532***	2.955***	2.922***	2.442***	2.969***	2.327***
	(0.143)	(0.095)	(0.218)	(0.130)	(0.285)	(0.175)
School below cutoff	−5.818***	−3.307***	−3.127***	−1.995***	−3.127***	−1.901***
	(0.216)	(0.102)	(0.346)	(0.169)	(0.432)	(0.219)
School belonging to subsystem above cutoff			1.157***	1.354***	1.253***	1.554***
			(0.164)	(0.123)	(0.198)	(0.151)
School belonging to subsystem below cutoff			−0.989***	−0.896***	−1.059***	−1.012***
			(0.229)	(0.149)	(0.280)	(0.187)
Distance from school above cutoff (km)					−0.085***	−0.098***
					(0.016)	(0.012)
Distance from school below cutoff (km)					0.091***	0.088***
					(0.016)	(0.012)
Observations	363191	363191	181650	181650	181650	181650

Note: Estimates are of the (negative) ratio of the coefficient on the interaction between the school characteristic × (score ≥ cutoff) dummy and the distance to school coefficient (measured in km). Coefficient estimates are from a conditional logit model for students within 10 points of an admission cutoff. Observations are weighted with respect to centered COMIPEMS score, using the edge kernel. Exploded logit estimates are for the student's top three choices. All specifications include distance to school, mean COMIPEMS score of students assigned to school, proportion of older sibling's middle school cohort choosing school, and school subsystem fixed effects. Also included are the uninteracted covariates for which the WTT estimates are reported, as well as interactions of these variables with first-order piecewise polynomials of centered COMIPEMS score. Standard errors accounting for clustering at the older sibling level are in parentheses. * p<0.10, ** p<0.05, *** p<0.01.

Figures

Figure 1: Distribution of cutoff scores (2011 exam year)

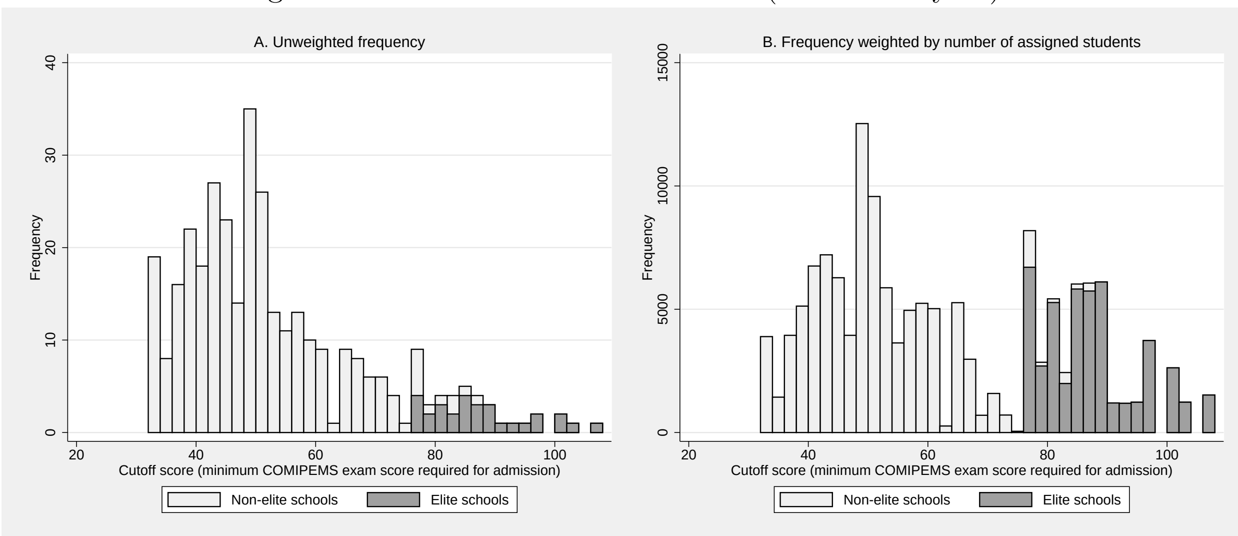
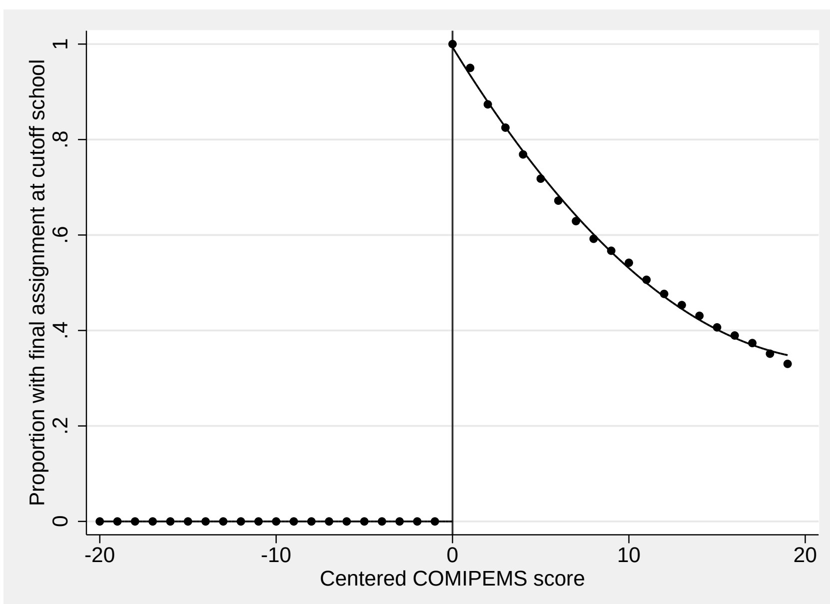
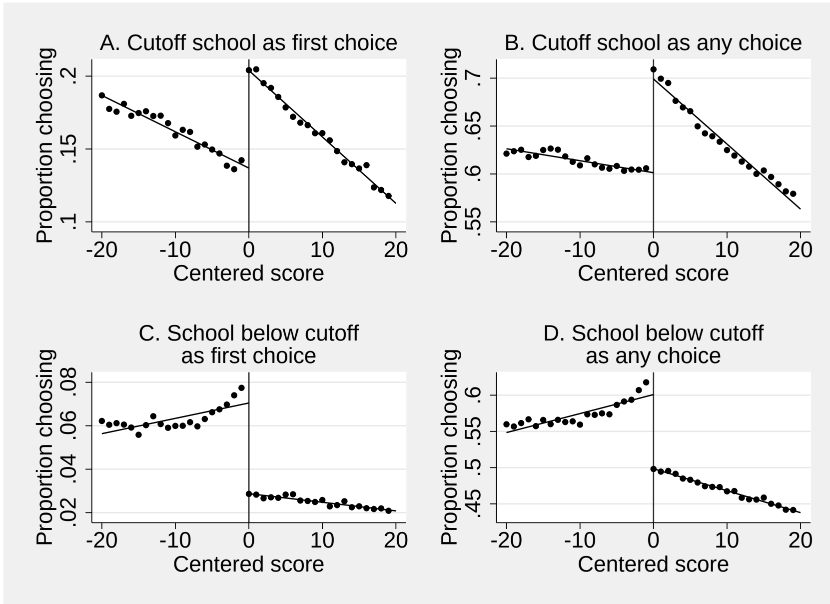


Figure 2: Verification of sharp discontinuity in admission probability due to assignment rule



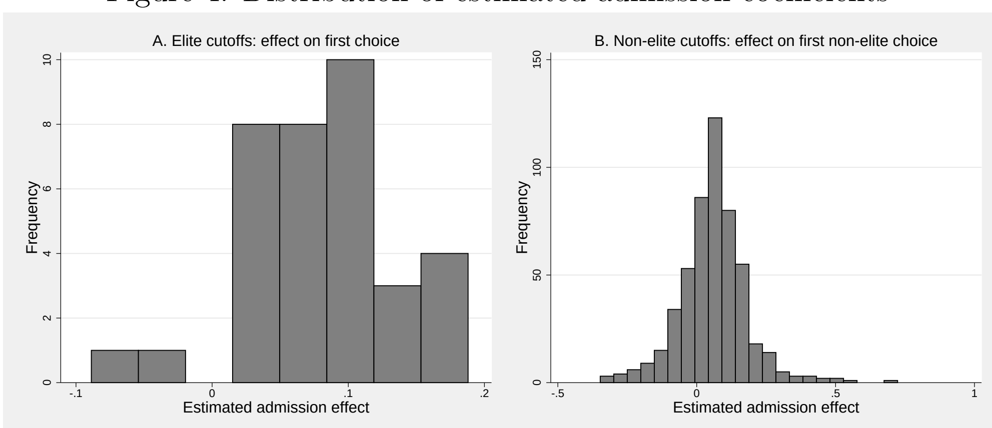
Note. Variable on vertical axis is proportion of older siblings assigned to the cutoff school. Variable on horizontal axis is older sibling's COMIPEMS exam score, centered to be 0 at the corresponding cutoff score.

Figure 3: Effect of older sibling admission on younger sibling choice



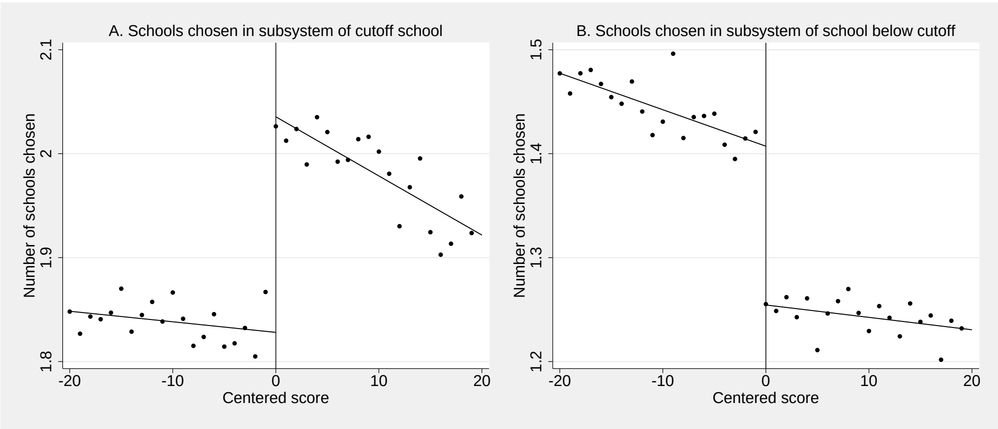
Note. Proportion on vertical axis pertains to the younger siblings. Variable on horizontal axis is older sibling's COMIPEMS exam score, centered to be 0 at the corresponding cutoff score. Fitted lines are from a linear fit.

Figure 4: Distribution of estimated admission coefficients



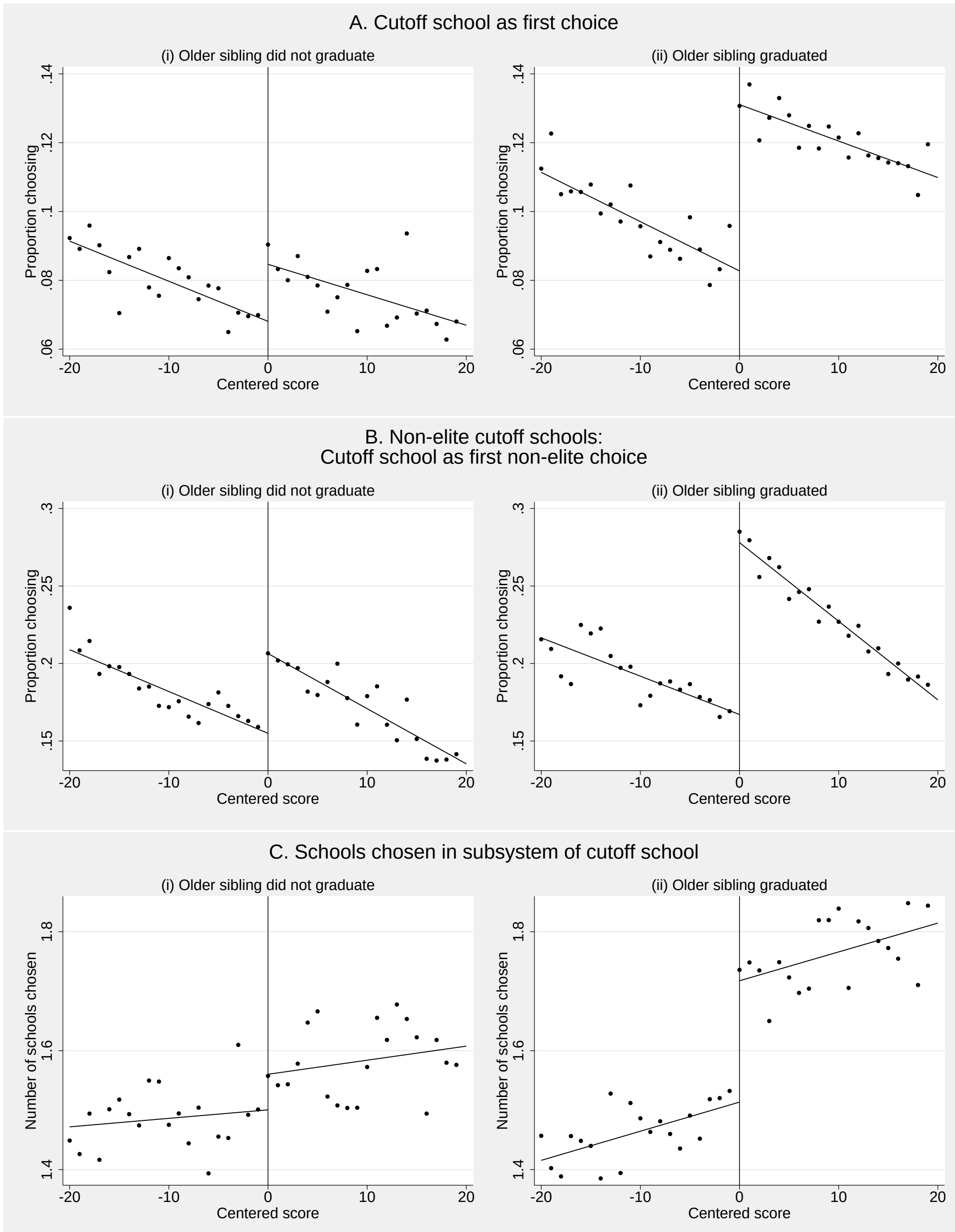
Note. Histograms are of estimated coefficients on older sibling admission, estimated from separate local linear regressions for each cutoff school using the IK optimal bandwidth, where the dependent variable is a dummy variable equal to 1 if the younger sibling chose the cutoff school as his first choice. Panel A plots the coefficients for elite cutoff schools. Panel B plots the coefficients for non-elite cutoff schools.

Figure 5: Effect of older sibling admission on number of other schools chosen in cutoff subsystem



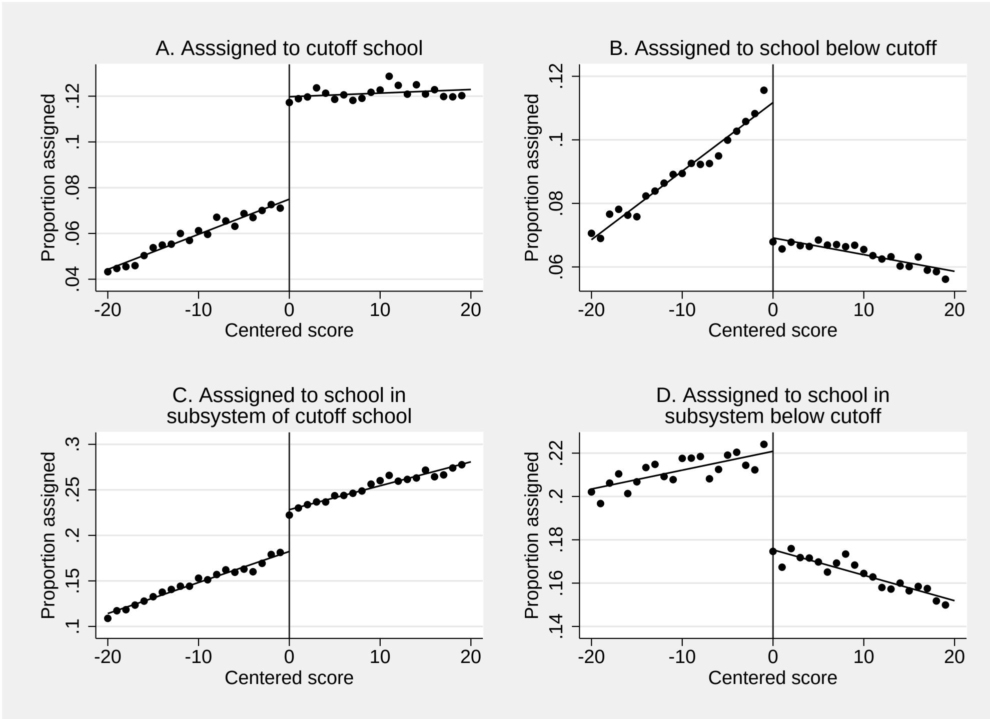
Note. Variable on vertical axis is number of schools selected by the younger sibling in the subsystem to which the older sibling's cutoff school belongs, excluding the cutoff school. Sample is limited to older siblings for whom the school immediately below the cutoff belongs to a different subsystem than the cutoff school, so that marginal rejection results in assignment to another subsystem. Variable on horizontal axis is older sibling's COMIPEMS exam score, centered to be 0 at the corresponding cutoff score. Fitted lines are from a linear fit.

Figure 6: Effect of older sibling admission, by graduation outcome



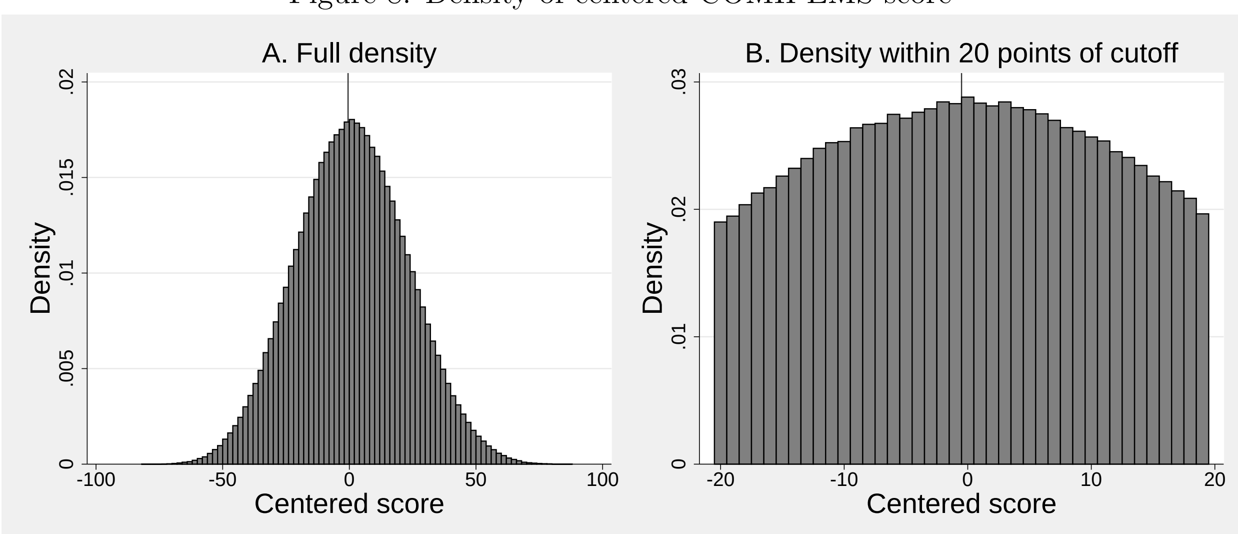
Note. Vertical axis pertains to the choice of the younger sibling. Number of schools chosen in Panel C excludes the cutoff school. Sample in Panel C is limited to older siblings for whom the school immediately below the cutoff belongs to a different subsystem than the cutoff school, so that marginal rejection results in assignment to another subsystem. "Graduation" is proxied by an indicator of whether the older sibling took the 12$^{\text{th}}$ grade standardized exam. Variable on horizontal axis is older sibling's COMIPEMS exam score, centered to be 0 at the corresponding cutoff score. Fitted lines are from a linear fit.

Figure 7: Effect of older sibling admission on younger sibling assignment outcomes



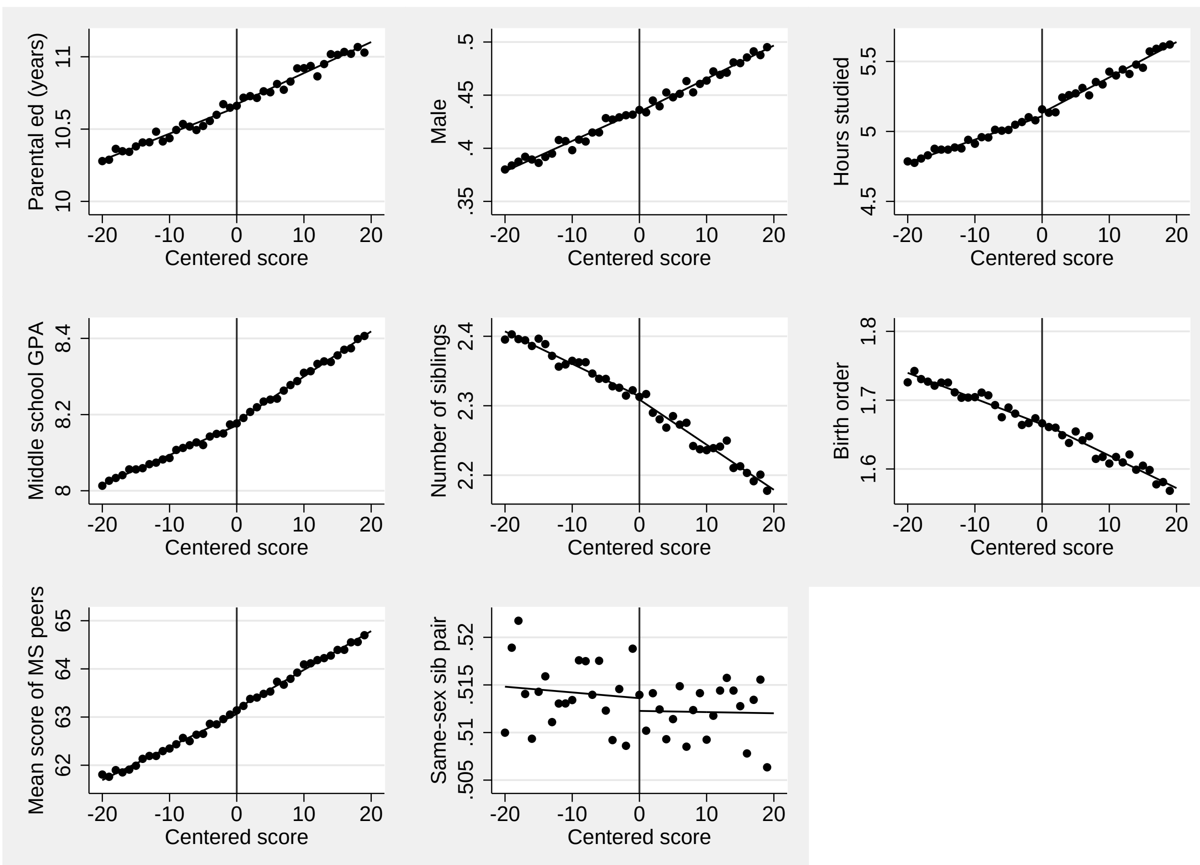
Note. Proportion on vertical axis pertains to the assignment outcome of the younger sibling. Variable on horizontal axis is older sibling's COMIPEMS exam score, centered to be 0 at the corresponding cutoff score. Fitted lines are from a linear fit.

Figure 8: Density of centered COMIPEMS score



Note. Histogram is of COMIPEMS score for students near a cutoff. Scores are centered so that they are 0 at the cutoff score.

Figure 9: Balance of exogenous covariates with respect to admission



Note. Dependent variables indicated on y-axis. Fitted lines are from a linear fit.