ANNENBERG
BROWN UNIVERSITY

EdWorkingPaper No. 19-165

O Brother, Where Start Thou? Sibling Spillovers in College Enrollment

Joshua Goodman
Brandeis University

Michael Hurwitz
College Board

Christine Mulhern
Harvard University

Jonathan Smith
Georgia State University

We study within-family spillovers in college enrollment to show college-going behavior is transmissible between peers. Because siblings' test scores are weakly correlated, we exploit college-specific admissions thresholds that directly affect older but not younger siblings' college options. Older siblings' admissibility substantially increases their own four-year college enrollment rate and quality of college attended. Their improved college choices in turn raise younger siblings' college enrollment rate and quality of college chosen, particularly for families with low predicted probabilities of college enrollment. Some younger siblings follow their older sibling to the same campus but many upgrade by choosing other colleges. The observed spillovers are not well-explained by price, income, proximity or legacy effects, but are most consistent with older siblings transmitting otherwise unavailable information about the college experience and its potential returns. The importance of such personally salient information may partly explain persistent differences in college-going rates by income, geography and other characteristics that define a community.

VERSION: December 2019

O Brother, Where Start Thou?
Sibling Spillovers in College Enrollment*

Joshua Goodman
Brandeis University
JoshuaGoodman@brandeis.edu

Michael Hurwitz
College Board
mhurwitz@collegeboard.org

Christine Mulhern
Harvard University
mulhern@g.harvard.edu

Jonathan Smith
Georgia State University
jsmith500@gsu.edu

November 19, 2019

Abstract

We study within-family spillovers in college enrollment to show college-going behavior is transmissible between peers. Because siblings' test scores are weakly correlated, we exploit college-specific admissions thresholds that directly affect older but not younger siblings' college options. Older siblings' admissibility substantially increases their own four-year college enrollment rate and quality of college attended. Their improved college choices in turn raise younger siblings' college enrollment rate and quality of college chosen, particularly for families with low predicted probabilities of college enrollment. Some younger siblings follow their older sibling to the same campus but many upgrade by choosing other colleges. The observed spillovers are not well-explained by price, income, proximity or legacy effects, but are most consistent with older siblings transmitting otherwise unavailable information about the college experience and its potential returns. The importance of such personally salient information may partly explain persistent differences in college-going rates by income, geography and other characteristics that define a community.

*We are grateful to the College Board for sharing its data for this project. For helpful feedback, we thank seminar and conference participants at the University of Chicago, Harvard, UVA, Brandeis, Michigan State, APPAM and AEFP. Melanie Rucinski provided excellent research assistance. Christine Mulhern was supported, in part, by the Institute of Education Sciences, U.S. Department of Education, through grant R305B150010 for the Partnering in Education Research Fellowship in collaboration with the Center for Education Policy Research at Harvard University. The opinions expressed are those of the authors and do not represent the views of the Institute or the U.S. Department of Education. All errors are due to our siblings.

1 Introduction

The college enrollment patterns of US students vary substantially by income and location, as well as other dimensions. Nationally, students from the highest income one percent of families are 77 times more likely to attend an Ivy League college than those from the bottom quintile (Chetty et al., 2017). Even among similarly low income students, enrollment rates vary substantially by geography. For those in the 25th percentile of the local parental income distribution, college enrollment rates range from less than 32 percent in the lowest-attending decile of commuting zones to over 55 percent in the highest decile commuting zones (Chetty et al., 2014).[1] Some of the most commonly studied explanations for such gaps by income and geography include credit constraints (Belley and Lochner, 2007; Lochner and Monge-Naranjo, 2012), differences in teacher and school quality (Chetty et al., 2014; Deming et al., 2014), and spatial variation in proximity of college options (Hillman, 2016).

Economists have paid less attention to social factors common to members of a given community that may partly explain such income-based and geographic differences in college enrollment patterns. If students' college choices are influenced by the choices of their peers, such spillovers would contribute to observed enrollment differences. High income students might be more likely than low income students to enroll at all or to choose high quality colleges in part because of exposure to other high income peers who have made such choices. Communities with historically low college enrollment rates may struggle to improve such rates because current students have few peers with firsthand experience of the enrollment process and life on a college campus. Such spillovers could partly explain persistence in college-going "culture" across communities.

Suggestive evidence that such social factors matter for college enrollment is scattered throughout recent papers. Hoxby and Avery (2013) note that high-achieving, low income students who fail to apply to any selective college are disproportionately from schools where such students "have only a negligible probability of meeting a... schoolmate from an older cohort who herself attended a selective college." Dillon and Smith (2017) show that, controlling for a rich set of covariates, the share of a student's high school graduates enrolling in four-year college strongly predicts the

[1]See Online Appendix Figure VII, panel B.

quality of college chosen by that student, perhaps because they "have many role models to follow through the college choice process." Black et al. (2015) similarly find that Black and Hispanic students in Texas are more likely to apply to colleges that "recently enrolled any students from the same high school and recently graduated students from the same high school of the same race". They hypothesize that students have better information about a given campus when close peers have attended that campus, and that successful degree completion by such peers is a particularly important signal. Among high-achieving students in the UK, Campbell et al. (2019) find that high schools' college enrollment rates statistically account for about half of the socioeconomic gap in quality of college chosen.

Siblings are particularly important peers. When asked who most influences their thinking about postsecondary education, half of American high schoolers choose family members, compared to only four percent who choose friends (Oymak, 2018). Numerous studies suggest that siblings are particularly influential and even moreso when parents lack college education. These include smaller scale qualitative studies showing that Black and Hispanic students report older siblings as important influences (Mwangi, 2015), as well as larger scale quantitative work showing that Black students' college enrollment rates are much higher when older siblings have enrolled earlier (Loury, 2004). Our prior work using data on 1.6 million sibling pairs of SAT-takers shows siblings' college choices are very closely related. One-fifth of younger siblings enroll in the same college as their older sibling and younger siblings are 15–20 percentage points more likely to enroll in four-year or highly competitive colleges if their older siblings do so first (Goodman et al., 2015). Correlations between sibling choices remain even after inclusion of extensive controls for potential confounds, such as siblings' academic achievement.

Such descriptive evidence does not, however, prove that older siblings' college choices causally influence the choices of their younger siblings. The empirical challenges of identifying such peer effects are well-known (Manski, 1993; Angrist, 2014). If two peers make similar college choices, two major issues arise in estimating peer effects. First, the "reflection problem" involves difficulty in distinguishing whether the first peer affects the second one or vice versa. We solve the reflection problem by estimating spillover effects from older to younger siblings, making the reasonable assumption that only the former's college choice can influence the latter's.[2] Second, the "common shock problem" arises from the fact that peers tend to share characteristics or environments that might be driving both of their choices. This problem is particularly acute when studying siblings, who share parents, homes, neighborhoods and other important potential determinants of college choice.

We solve the common shock problem by finding exogenous variation that affects the colleges available to older siblings but does not directly affect younger siblings through any channel other than their older siblings' college choice. The opening of a nearby college, used for example in Currie and Moretti (2003), would not work as such an instrument because it would directly affect the college options of both older and younger siblings. Instead, using data on the universe of SAT-takers with no siblings, we identify a set of colleges that appear to use hidden admissions thresholds.[3] We then identify older siblings who applied to such colleges, which we refer to as the older siblings' "target" colleges. A regression discontinuity design using older siblings' distance to the relevant threshold provides clear identification of such thresholds' impacts on older siblings' choices. Because younger siblings' SAT scores are only weakly correlated with older siblings' scores, the thresholds do not directly affect younger siblings' own admissions outcomes. Any observed impact on younger siblings' college choices must thus run through older siblings' choices.

Our first stage analysis shows that meeting these admissions thresholds improves older siblings' college choices. Admissibility substantially increases older siblings' likelihood of enrolling in the target college, which has higher graduation rates and peer quality than the four- or two-year institutions they otherwise would attend. For the older sibling, admissibility provides exogenous variation in target college enrollment, which in turn thus substantially increases older siblings' four-year college enrollment rate and quality of college attended.

Such increases in older siblings' college enrollment and quality spill over onto younger siblings' college choices. Older siblings' enrollment in the target college induces some younger sib-

[2]See Black et al. (2017) for a similar discussion of identifying peer effects among siblings.

[3]We do not observe admissions decisions but infer these thresholds from discontinuities in enrollment rates. Such patterns could in theory also be driven by thresholds that affect financial aid packages, which in turn affect enrollment probabilities.

lings to follow them to that same college. It induces a similar number of younger siblings to attend other four-year colleges. Many would not otherwise have attended college at all, so that younger siblings' college enrollment rates and quality of chosen colleges increase substantially. Importantly, spillover effects on college enrollment and quality are much larger for those whose family characteristics predict lower rates of enrollment in four-year colleges.

What explains sibling spillovers in college enrollment? We consider and largely reject four potential channels as insufficient to explain the observed patterns. Income effects, whereby family credit constraints imply one sibling's college spending can crowd out another sibling's, are inconsistent with our results showing older siblings' slightly costlier college choices induce younger siblings to make costlier choices as well. Price effects, whereby a younger sibling receives more financial aid because an older sibling is also attending college, do not explain why we observe spillovers for siblings far enough apart in age as to not be enrolled simultaneously. Proximity effects, whereby siblings benefit (or suffer) from being enrolled on the same campus, can not explain why many younger siblings are induced to attend different colleges than the older sibling. Legacy effects, whereby colleges give admissions preferences to those whose family members previously attended, similarly can not explain why many younger siblings attend different colleges than their older siblings.

We conclude that the explanation most consistent with the observed spillovers involves information effects. Older siblings' enrollment in higher quality colleges generates for parents and younger siblings information unlikely to be transmitted through other means. Older siblings attending higher quality colleges than they otherwise would may report to their families more positive experiences from those campuses. Higher persistence and graduation rates on such campuses may, in particular, change older siblings' reports of the returns to college enrollment. Such information from older siblings may change parents or younger siblings' perceptions of younger siblings' potential returns to attending higher quality colleges. Younger siblings may place particularly high weight on their older siblings' college experiences, given that the educational success of a close family member may be more salient and predictive of one's own success than less personalized sources of information.

These results contribute to two additional major strands of recent research. First, we provide some of the only evidence, and the only in the US context, of peer effects in college choices. Until recently, most of the voluminous peer effects literature exploited random or quasi-random assignment of classmates, schoolmates or roommates to study spillovers of peers' characteristics or risky behaviors onto students' own academic achievement or risky behaviors (Sacerdote, 2011). That literature rarely, if ever, focused on siblings as peers or considered college choices as either treatments or outcomes. Recent research has begun to provide evidence of spillovers between siblings in various behaviors, including: smoking and drinking (Altonji et al., 2017); military service (Bingley et al., 2019); and paternity leave usage (Dahl et al., 2014). The latter two papers argue that increased information, about the returns to military service and employers' reaction to leave-taking, are the most likely mechanism explaining sibling spillovers in these non-educational choices.

A handful of recent papers from outside the US suggest sibling spillovers in educational choices. Using distance to the nearest girls' school as an instrument, Qureshi (2018) shows that additional schooling for Pakistani eldest sisters induces younger brothers to pursue more schooling. Joensen and Nielsen (2018) use quasi-random variation in a school pilot scheme to show Danish older siblings' pursuit of advanced math and science coursework increases younger siblings' propensity to take such courses. Dustan (2018) uses randomness induced by Mexico City's high school assignment mechanism to show students prefer schools older siblings have attended. Most similar to this paper are two recent working papers exploiting discontinuities in Chilean, Croatian and Swedish admissions systems to estimate sibling spillovers in choice of degree and institutions (Aguirre and Matta, 2018; Altmejd et al., 2019). Those papers also find spillovers across siblings too far apart in age to be enrolling simultaneously and suggest that information transmission is the most plausible explanation for the observed effects.

A second related literature, consistent with our argument that siblings can provide important information too costly or impossible to obtain otherwise, argues that "low-touch" interventions substantially underperform "high-touch" ones with respect to college choice. Multiple recent papers using nudge-style informational interventions at state or national scale fail to meaningfully impact college enrollment choices (Gurantz et al., 2019; Bird et al., 2019; Hyman, 2019). Researchers testing multiple treatments often find that information only interventions have little impact on students (Bettinger et al., 2012; Carrell and Sacerdote, 2017). Information on college-level earnings, released through the US federal government's College Scorecard website, does not affect the college application patterns of students from non-wealthy families or high schools (Hurwitz and Smith, 2018). One explanation consistent with these findings is that students, particularly low income ones, may not view as salient the information provided by such interventions, information which is rarely personalized to their own circumstances.

Conversely, higher touch interventions with a more personalized flavor seem to make substantial differences in the college choices of low income students. Carrell and Sacerdote (2017), for example, show in-person mentoring improves college attendance, perhaps by substituting for costly or absent time investments of parents and teachers. Intensive after-school college counseling can improve the quality of colleges first generation students choose (Castleman and Goodman, 2018; Barr and Castleman, 2018). Traditional high school guidance counselors can make a substantial difference in the college choices of disadvantaged students, in part through the provision of information (Mulhern, 2019a). College choices can even be affected by personalized software showing students how their admission chances compare to prior students from their own high school with similar GPAs and SAT scores (Mulhern, 2019b). This body of work suggests students react powerfully to information from people close to them or sources that otherwise are highly personalized and thus salient. Siblings, and socially close peers more generally, are one critical source of such information.

2 Data and Empirical Strategy

2.1 Data

We begin with student-level data from the College Board on all students from the high school classes of 2004-14 who took the PSAT, the SAT, or any Advanced Placement exams. We observe each student's name, home address and high school attended, as well as self-reported demographic information on gender, race, parental education and family income. We also observe scores from all takes of the SAT, which during this time is offered seven times a year and can be taken as often as the college application timeline allows.[4] We observe all section scores from all takes, allowing us to construct the SAT superscores most commonly used by college admissions processes (Goodman et al., 2020). The superscores are the sum of a student's maximum math and maximum critical reading scores, regardless of whether those scores occurred on the same take.

For college outcomes, we observe all colleges to which students send their SAT scores, which serves as a decent proxy for college application behavior (Smith, 2018). A merge between the College Board data and the National Student Clearinghouse (NSC), which tracks the enrollment of students at nearly all U.S. colleges, allows us to observe the specific college a student enrolls in at any given moment in time. We obtain some characteristics of such colleges, such as the average net price paid by students, by merging the NSC data to the federal government's Integrated Postsecondary Education Data System (IPEDS).

We construct two measures of college quality because neither NSC nor IPEDS provide useful measures that are comparable across all colleges. First, we use the NSC data to construct for each college the fraction of initial enrollees who eventually earn a B.A. from any college within six years. Relative to IPEDS' graduation rate measures, this has the advantage of accounting for transfers between institutions and allows direction comparisoon of two- and four-year colleges. Students who do not enroll in college are assigned a zero for this variable. Second, we follow Smith and Stange (2016) and construct for each college the average standardized PSAT score of initial enrollees. This peer quality measure again has the advantage of allowing comparisons between two- and four-year institutions, the former of which do not require SAT scores and thus lack such a peer quality measure in IPEDS. Students who do not enroll in college are assigned the mean PSAT score of all such students for this variable.

[4] During this time period, retakes cost roughly $40, with low income students eligible for fee waivers for up to two takes.

2.2 Identifying Siblings

We identify siblings as pairs of students from different high school classes whose last name and home address match. This approach should yield relatively few false positives, such as cousins living together. It does, however, likely generate many false negatives in which we mistakenly label individuals with siblings as only children. This can stem from two primary sources. First, we fail to identify as siblings any students whose family changes residential addresses in between the years in which those siblings took the SATs. Second, and likely a smaller concern, siblings may record their last names or home address differently.[5]

That we fail to identify such students as having siblings will have no impact on the internal validity of our subsequent estimates but does affect both sample size and the characteristics of the population we study. In order to have a sibling in this data, families must have at least two children in the high school classes of 2004-14, each of whom takes at least one College Board test and who live at the same home address at the last test take. We refer to anyone for whom we fail to identify a sibling as an "only child".

Because some families have more than two siblings, we use each family's oldest sibling to determine the "treatment" status of all younger siblings. The vast majority of siblings in our data (77 percent) appear in pairs but 23 percent come from families where identify three or more siblings. Because we use SAT-based admissions thresholds for variation in college access, our sample therefore consists of families where the oldest identified sibling took the SAT. Families' demographic characteristics are assigned to all younger siblings based on oldest siblings' reports, for consistency across siblings and because treatment status is determined at the time of oldest siblings' college applications. We structure the data so that each observation is a younger sibling, whose characteristics and treatment status are assigned based on their oldest sibling.

Our starting samples, shown in the first two columns of Table 1, consist of 11.3 million SAT-takers with no identifiable sibling in the data and 3.1 million younger siblings with an identifiable oldest sibling who took the SAT. Panel A summarizes demographic characteristics common to

[5]Our matching process also identifies twins as only children because they are in the same high school class. We do this in order to generate a set of siblings where influences clearly run from older to younger siblings. With twins, the direction of influence is unclear.

all siblings in a family, while panels B and C show SAT scores and college outcomes of older and younger siblings separately. Because siblings are more likely to be identified in residentially stable families and those where test-taking is more common, the sibling sample is more economically and educationally advantaged than the only child sample. Relative to only children, siblings have higher family income and more educated mothers, are more likely to be white, and have higher SAT scores and college enrollment rates. The sibling sample is, however, somewhat diverse, with 20 percent coming from black or Hispanic families and 35 percent having mothers who did not attend college.

2.3 Detection of Admissions Thresholds

We use our data to detect which colleges admit students in part on the basis of minimum SAT thresholds not known to applicants. Many colleges use minimum SAT scores as one criterion for determining admissions decisions, so that meeting or exceeding a college's threshold typically increases a student's probability of being admitted to that college. We focus on thresholds hidden from applicants because publicly known thresholds induce some students to retake the SAT until their scores meet the thresholds (Goodman et al., 2017). Such behavior creates endogenous sorting around the threshold that invalidates the regression discontinuity design.

Conversely, students can not react endogenously to cutoffs about which they are unaware. Colleges that use such thresholds but not make this fact public are, for obvious reasons, unlikely to share information about such practices with researchers. We therefore search our data for colleges that use minimum SAT thresholds in the admissions process but do not make this fact public.

We search for such thresholds using the only child sample, which is independent of the sibling sample we use to estimate spillover effects. This avoids the potentially spurious findings that might be generated by searching for thresholds using the same observations and outcomes used to estimate treatment effects. For each college and year, we find all only children who sent their SAT scores to that college, generating an indicator for a student enrolling in that college within one year of graduating high school. We then search for discontinuities by SAT score in a given college's enrollment rate among its applicants. We limit our search to the 526 colleges that received SAT scores from at least 1,000 students each year, in order to minimize the possibility of false positives arising from small samples.

To search for discontinuities, we estimate local linear regression discontinuity models at each SAT score that might represent a potential threshold.[6] We define the set of potential thresholds for each college as the set of SAT scores in the 5th to 50th percentiles of the applicant distribution for the specified college and year. Colleges are unlikely to set minimum thresholds lower or higher in their applicant distributions. For every potential threshold T and all applicants i to college c in year y, we run regressions of the form:

$$Enrolled_{icy} = \beta_0 + \beta_1 \mathbb{1}[SAT_i \geq T_{cy}] + \beta_2 (SAT_i - T_{cy}) + \beta_3 \mathbb{1}[SAT_i \geq T_{cy}] \times (SAT_i - T_{cy}) + \epsilon_{icy} \tag{1}$$

We define the running variable using students' SAT superscores, the most frequently used form of scores considered by college admissions offices. To minimize false positives driven by specification error, we use a bandwidth of 60 SAT points, within which enrollment graphs look generally linear.

The coefficient of interest β_1 estimates the magnitude of any potential discontinuity in enrollment rates at the given threshold T. To further limit potential false positives, we consider as discontinuities only those instances where discontinuities in enrollment rates exceed five percentage points and where we reject the null hypothesis of no discontinuity with $p > 0.0001$. Finally, we discard any colleges where thresholds are detected in fewer than five years, given that most colleges that use minimum SAT scores in admissions are unlikely to change that policy from year to year. We also discard a small number of colleges for which we find evidence from admissions websites that the detected thresholds are publicly known.

This procedure yields 21 threshold-using colleges, which we refer to as "target" colleges both for brevity and because of older siblings' interest in attending these institutions. These target colleges are largely public institutions (16 public, 5 private) with an average enrollment of over 10,000 full-time equivalent students, and are located in eight different East coast states. The median SAT

[6] Our approach is similar to that used in Andrews et al. (2017).

threshold across years for these colleges ranges from 720 to 1060, with students relatively widely distributed across these colleges and thresholds.[7] These target colleges' average graduation rate is 63 percent and the average PSAT z-score of their students is 0.27. They have average net prices of $12,500, making them $4,000 less expensive per year than the average college attended by students in our full sample.

Our last use of the only child sample is to predict college enrollment rates in our sibling sample based on families' demographic characteristics and college applications. Using only children who applied to one of the target colleges, we regress a four-year college enrollment indicator on indicators for all observed categories of income, race, maternal education and paternal education, as well as indicators for the target college applied to. This has the advantage of using all demographic information simultaneously and dealing with missing information by treating such missingness as its own category.[8] We then use the coefficients from that regression to make out-of-sample predictions of four-year college enrollment rates in our sibling sample.

We label the bottom third of the predicted enrollment distribution as "uncertain college-goers" and the top two-thirds as "probable college-goers". Uncertain college-goers have a predicted four-year college enrollment rate of 62 percent, compared to probable college-goers' predicted enrollment rate of 77 percent. These rates are fairly high because the sample is already conditioned on families with oldest siblings who have taken the SAT and applied to four-year colleges. Nonetheless, there is still sufficient variation in these predicted enrollment rates to capture meaningful differences in families' attachment to and engagement with postsecondary education.

Having used the only child sample to identify target colleges and make enrollment predictions, we now set the only child sample aside and perform all subsequent analysis on the sibling sample. Our main analytical sample is the set of younger siblings whose oldest sibling sent SAT scores to a target college in a year when that college was using an admissions threshold. This yields over 166,000 younger siblings, 77 percent of whom have only one older sibling in the data and 23 percent of whom have at least two siblings in the data. The sample of younger siblings whose

[7] Figure 1 shows the number of younger siblings in each college represented in our final regression discontinuity sample, described further below.

[8] Among our final sibling sample, fewer than one percent are missing race, about 10 percent are missing parental education, and 35 percent are missing income.

oldest siblings applied to target colleges looks fairly similar demographically to the full sibling sample, as seen in column 3 of Table 1. Both are 66 percent white, have average family income around $80,000, and have oldest siblings' average SAT scores somewhat under 1100.

2.4 Estimation of Sibling Spillovers

Identification of spillover effects comes from comparison of families whose oldest siblings applied to the same target college in the same year with nearly identical SAT scores, but whose college options differed only because of a hidden admissions threshold. We first characterize the treatment by exploring how target colleges' admissions thresholds changed older siblings' college choices. Our first stage regression is a local linear regression discontinuity design of the form:

$$OlderTarget_{ifcy} = \alpha_0 + \alpha_1 OlderAccess_f + \alpha_2 Distance_f + \alpha_3 OlderAccess_f \times Distance_f + \lambda_{cy} + \epsilon_{ifcy}. \quad (2)$$

Here, $OlderTarget$ indicates older siblings' target college enrollment, with subscripts denoting younger sibling i, from family f, whose oldest sibling applied to target college c in year y. $OlderAccess$ indicates the older sibling's SAT superscore met or exceeded the admissions threshold and $Distance$ represents the older sibling's distance from that threshold. The coefficient of interest α_1 estimates the difference in target college enrollment rates between older siblings just above and below the threshold. Target college-by-year fixed effects λ imply that all such comparisons are made only between older siblings applying simultaneously to the same target college.

To study the impact of older siblings' college access on their own college choices and those of their younger siblings, we could then estimate reduced form regressions of the form:

$$Y_{ifcy} = \gamma_0 + \gamma_1 OlderAccess_f + \gamma_2 Distance_f + \gamma_3 OlderAccess_f \times Distance_f + \lambda_{cy} + \epsilon_{ifcy}, \quad (3)$$

where Y characterizes the college choices of older or younger siblings.

We instead focus on instrumental variables estimates of the impact of older siblings' target college enrollment on younger siblings' college choices. We do so in part because we do not observe actual admissions and thus can not scale any reduced form impacts in a meaningful way. The instrumental variables approach also makes what we believe is a warranted assumption that older siblings' access to the target college can affect younger siblings only through older siblings' enrollment in that college. In other words, these estimates assume that older siblings' admission alone can not affect younger siblings unless the former actually enroll in the target college.

We therefore run regressions of the form

$$Y_{ifcy} = \beta_0 + \beta_1 OlderTarget_f + \beta_2 Distance_f + \beta_3 OlderAccess_f \times Distance_f + \lambda_{cy} + \epsilon_{ifcy}, \quad (4)$$

where older siblings' target college enrollment $OlderTarget$ is instrumented by their threshold-driven access, using first stage Equation 2. We first describe the "treatment" more fully by defining outcomes Y that characterize older siblings' college choices, in order to show how target colleges differ from the counterfactual college choices older siblings would have made. We then study younger siblings' college choices, estimating spillovers onto younger siblings' target college enrollment, four- and two-year enrollment, quality of chosen college chosen, and price and location of chosen college.

Our default specification uses a bandwidth of 100 SAT points, roughly the (Imbens and Kalyanaraman, 2012) optimal bandwidth for many outcomes, though we show our main results are robust to different bandwidth choices. Because our hidden thresholds may be measured with noise, we also default to a donut hole specification that excludes observations on the threshold itself. We later show are results are robust, though slightly less precise, to inclusion of such observations. Finally, we cluster our standard errors by oldest siblings' high schools, which accounts for unobserved high school-specific correlations in prediction errors and nests within it clustering by family, to account for the fact that some older siblings are represented multiple times in the data across their multiple younger siblings.

Our final regression discontinuity sample, shown in column 4 of Table 1, contains over 49,000 younger siblings. Relative to the full sample of only children and siblings, our RD sample is more racially diverse but fairly similar in income and parental education. Columns 5 and 6 separately show the uncertain and probable college-goers from this RD sample. Only 45 percent of younger siblings in the former group enroll in four-year college, compared to 65 percent in the latter group, suggesting our predictive analysis successfully detected the relevant heterogeneity.

Before turning to our main results, we first confirm the validity of our regression discontinuity approach. The distribution of observations near the hidden admissions thresholds appears smooth.[9] Excess mass to the right of the threshold would imply that older siblings were either retaking the SAT to meet the admissions threshold or that those who had met the threshold were disproportionately likely to apply to that college in the first place. The absence of such excess mass suggests that students are unaware of the existence of these thresholds. Similarly, we see no evidence of imbalance in any of the observed demographic characteristics, either in the full sample or among uncertain and probable college-goers separately.[10] Younger siblings whose oldest siblings fall just above and below hidden admissions thresholds appear observationally identical. They appear to differ only in their oldest siblings' potential access to the threshold-using college.

That siblings' SAT scores are not very highly correlated rules out one final threat to validity, namely that target colleges' admissions thresholds are affecting younger siblings not through their older siblings' college options but directly through their own. Among the nearly 90 percent of younger siblings who took the SAT, the distribution of differences between their scores and those of their older siblings is 175 SAT points, or 165 points among the regression discontinuity sample).[11] This means that the set of older siblings just above and below the admissions thresholds rarely have younger siblings who are similarly near those thresholds. The threshold thus provides exogenous variation in older siblings' target college access distinct from younger siblings' access, making the exclusion restriction more plausible.

[9] See Figure A.1.
[10] See Figure A.2 and Table A.1.
[11] See Figure A.3.

3 Results

3.1 Older Siblings' College Choices

We begin our causal analysis by documenting the impact of hidden admissions thresholds on older siblings' college choices. Our goal is to fully characterize how such thresholds alter the types of colleges older siblings attend, in order to better understand the nature of the treatment younger siblings are being exposed to. We find that meeting these thresholds substantially increases older siblings' enrollment in target colleges and that such colleges are of much higher quality and, if anything, higher cost than the college choices older siblings would otherwise have made. We now discuss these findings in detail.

Threshold-based admissibility roughly doubles the probability that older siblings enroll in the target college. Figure 2 shows the remarkably linear first stage relationship between an older sibling's distance to the target college's admissions thresholds and his or her probability of enrolling in that target college. The visually striking discontinuity suggests that those just above the threshold are roughly 10 percentage points, or nearly twice as likely, to enroll in target colleges as those just below the threshold. That the point on the threshold is farthest from the fitted line suggests potential mis-identification of the precise minimum SAT score used by the target colleges' admissions processes, which motivates our decision to default to donut hole specifications.

Table 2 quantifies that first stage discontinuity, estimating that meeting the minimum threshold increases by nine percentage points the probability an older sibling enrolls in the threshold-using college. The instrument is quite strong, with F-statistics exceeding 150 in our full sample default specification and exceeding 20 even with narrower bandwidths and among smaller sub-samples. The point estimate is quite robust to different specifications, ranging only from 8-10 percentage points across bandwidths of 60-140 SAT points and remaining unchanged by the addition of demographic controls. Eliminating the donut hole leaves the instrument quite strong but shrinks the first stage magnitude to six percentage points. The magnitude of the first stage varies little by predicted college enrollment rate, with quite similar estimates for uncertain and probable college-goers.

Nearly half of the marginal older siblings induced by admissions thresholds to attend target colleges would not have attended four-year colleges otherwise. Panel A of Figure 3 shows a clear discontinuity at the admissions threshold in the probability an older sibling enrolls in any four-year college. Instrumental variables estimates in Table 3 imply that enrolling in the target college increases older siblings' four-year college enrollment rate by 42 percentage points, from a baseline rate of 58 percent. Many of these older siblings would otherwise have attended two-year colleges, so that the two-year enrollment rate of marginal older siblings drops by 34 percentage points (to zero). Only eight percent of these marginal older siblings would not otherwise have attended college at all, so that access to the target college largely creates opportunities to choose four-year institutions over two-year ones. That pattern appears both for uncertain and probable college-goers.

For marginal older siblings, target colleges are of substantially higher quality than the two-year and four-year alternatives they would otherwise have chosen. Panel B of Figure 3 shows a discontinuity at the admissions threshold in the historical B.A. completion rate of older siblings' chosen colleges. Instrumental variables estimates in Table 3 imply that target colleges have B.A. completion rates 26 percentage points higher than the 41 percent completion rate older siblings' counterfactual colleges would have had. Peer quality at target colleges, as measured by student body mean PSAT scores, is 0.45 standard deviations higher than the -0.06 average at older siblings' counterfactual colleges. The magnitude of older siblings' college quality improvements is somewhat larger for uncertain college-goers than for probable college-goers, though such differences are not statistically significant.

Older siblings' target colleges are, if anything, more costly for their families and farther from home than their counterfactual college options. Estimates in column 5 of Table 3 imply that target colleges' annual net prices are over $3,000 higher than the alternative, an unsurprising result given that two-year colleges generally have substantially lower tuition than do four-year colleges. This difference, though statistically insignificant, suggests that older siblings' target college enrollment may result in families having less savings to spend on younger siblings potential college tuition. Target college enrollment also substantially increases the likelihood that older siblings enroll in an institution at least 50 miles away from their home. Though not shown here, we observe weaker evidence that target college enrollment increases out-of-state enrollment, suggesting that most older siblings are enrolling in in-state colleges farther from home than they would otherwise have chosen.

Our estimates about the impact of hidden admissions thresholds on older siblings' college choices are robust to alternative specifications. The magnitude and statistical significance of our first stage and instrumental estimates are stable across all but the narrowest bandwidths between 60 and 140 SAT points, both for the full sample and for the sub-sample of uncertain college-goers.[12] The estimated discontinuities are largest at the true admissions thresholds and diminish when we use our default specification to test placebo thresholds at other points in the SAT score distribution.[13] This suggests our measured effects are true causal impacts and not simply statistical artifacts.

In summary, the admissions thresholds we identify serve as a strong instrument, generating substantial variation in older siblings' college choices. Admissibility increases older siblings' enrollment at target colleges. Many would otherwise have attended two-year colleges, so that this treatment involves large increases in four-year college enrollment and college quality, as measured by graduation rates and peers' academic skills. Older siblings end up at institutions that, if anything, cost their families more tuition than alternatives and end up farther from home.

3.2 Younger Siblings' College Choices

We turn to estimating the impact on younger siblings of having older siblings who attend colleges with higher BA completion rates and peer quality, and perhaps higher cost, than they otherwise would have. We first consider younger siblings' college application process and then consider their college enrollment decisions. We find that older siblings' enrollment in target colleges increases younger siblings' probability of applying to the target college, generally improves the quality of colleges they apply to, and substantially increases younger siblings' college enrollment rate and quality of college chosen. The impact of older siblings' college choices on younger

[12]For details, see Figures A.4 and A.5.
[13]For details, see Figure A.6.

siblings' college enrollment rate and college quality are particularly pronounced for uncertain college-goers, with probably college-goers seeing smaller spillovers. We now discuss these findings in detail.

Older siblings' enrollment in higher quality colleges does not affect their younger siblings' SAT-taking behaviors but does increase the quality of colleges to which they apply. The estimates in Table4 suggest that older siblings' target college enrollment has little clear impact on young siblings' probability of taking the SAT or on their ultimate SAT scores.[14] Older siblings' target college enrollment does, however, nearly quadruple the probability that younger siblings apply to the same target college (from 10 to 37 percent). Though younger siblings do not appear to increase their number of college applications, they roughly double their probability (from 25 to 53 percent) of applying only to colleges with historical graduation rates higher than 50 percent, suggesting their portfolio of applications improves given their older siblings choice of a higher quality college. Younger siblings' application patterns suggest interest in imitating older siblings' choices, both in terms of the specific college and general quality of college chosen.

Spillover effects result in some younger siblings following their older siblings to the target colleges to which the latter were exogenously granted access. Figure 4 shows the reduced form impact of older siblings' admissibility to the target college on their younger siblings' probability of enrolling in that same college. A large discontinuity in younger siblings' enrollment rate is clearly visible at the admissions threshold faced by the older sibling. As Table 5 shows, 13 percent of younger siblings follow their older sibling to the target college only because their older sibling enrolled there, an impact that is highly statistically significant. Our control complier estimates suggest that only two percent of younger siblings would otherwise have attended the target college, implying that nearly the only reason they attend is because their older sibling did. The extent to which younger siblings follow their older siblings to the same college is roughly twice as large for uncertain college-goers as for probable college-goers.

Older siblings' enrollment in higher quality target colleges increases the college enrollment

[14]To deal with potential selection into which younger siblings have valid SAT scores, we also create indicators for having a valid score above every multiple of 100. We see no clear evidence that older siblings' college choices affect any part of the SAT distribution for younger siblings, including the extensive margin.

rates of younger siblings, and not only due increased enrollment at the target colleges. The spillover effects are not simply the result of younger siblings substituting the target college for another four-year college they would otherwise have attended. Panel A of Figure 5 shows a clear discontinuity at older siblings' admissions thresholds in younger siblings' four-year college enrollment rate. As Table 5 shows, younger siblings four-year college enrollment rate increases by 28 percentage points, nearly doubling, with little of that increase representing substitution away from two-year colleges. Many younger siblings would not have attended college at all if their older sibling had not enrolled in the target college. That the coefficient in column two is twice the magnitude of the coefficient in column one implies that at least half the rise in four-year college enrollment rates comes from younger siblings choosing four-year colleges other than the target colleges themselves.

The target colleges and other four-year colleges that younger siblings enroll in are of higher quality than the counterfactual college choice they would have made. Panel B of Figure 5 shows a clear discontinuity at older siblings' admissions thresholds in the graduation rate of younger siblings' chosen colleges. Estimates in Table 5 imply that younger siblings' chosen colleges have graduation rates 20 percentage points higher and peer quality 0.35 standard deviations higher than the choices they otherwise would have made. Increases in younger siblings' college enrollment rates and college quality are driven almost entirely by uncertain college-goers, with much smaller spillover effects for probably college-going families.

Younger siblings also enroll in colleges that are, if anything, more costly than the choice they would have made had their older siblings not enrolled in the target college. On average, younger siblings choose colleges with net prices $3,600 higher than they otherwise would have, though that difference is not statistically significant. Uncertain college-goers drive these results, enrolling in colleges with net prices close to $10,000 higher. Given that older siblings are, if anything, increasing their families' costs by attending the target college, the fact that younger siblings' costs also rise implies families are spending more in total on their childrens' college educations. Any income effect therefore appears dominated by other channels and does not explain well the observed spillovers between siblings. Uncertain college-goers also appear more likely to choose colleges far from home as a result of their older siblings' enrollment in target colleges far from home.

Our estimates about the impact of older siblings' target college enrollment on younger siblings' college choices are robust to alternative specifications. The magnitude and statistical significance of our instrumental estimates are fairly stable across bandwidths between 60 and 140 SAT points, both for the full sample and for the sub-sample of uncertain college-goers.[15] The reduced form discontinuities peak at or close to the estimated admissions thresholds and diminish at placebo thresholds to the right and left, again suggesting our measured effects are not simply statistical artifacts.[16] The magnitude and statistical significance of these spillover effects are robust to inclusion of covariates in our default specification and generally robust to elimination of the donut hole, though the latter reduces their precision somewhat due to a diminished first stage.[17]

In summary, older siblings' college choices have large and clear spillover effects on the choices of their younger siblings. Younger siblings shift their college application portfolios to include the colleges that their older siblings attend and generally upgrade the quality of colleges they apply to because their older sibling enrolled in a higher quality college. Younger siblings then also enroll in higher quality colleges, in part by enrolling in four-year colleges when they would not have enrolled anywhere otherwise. These spillover effects are much larger for uncertain college-goers, those from families whose ex ante characteristics predict only moderate likelihood of attending four-year colleges at all.

3.3 Channels

What explains the college enrollment spillover effects from older to younger siblings? We consider five potential channels through which such spillovers might occur: income effects; price effects; proximity effects; legacy effects; and information effects. We discuss each of these in turn and conclude that only the last, information, is consistent with our results.

Income effects can generate spillovers if older siblings' college choices affect family budget constraints that determine younger siblings' college financing options. Older siblings' choices

[15]For details, see Figures A.7 and A.8.
[16]For details, see Figure A.6.
[17]For details, see Table A.2.

of more expensive colleges may leave credit constrained families with less funding to provide younger siblings. Our estimates suggest that older siblings' enrollment in target colleges slightly increases their cost of college. If income effects were large, we would then expect to see younger siblings making less costly college choices, including enrolling at lower rates. Instead, we observe younger siblings enrolling at higher rates, choosing colleges that are slightly more costly than they otherwise would have or much more costly in the case of uncertain college-goers. If income effects are present, they are dominated by other channels.

Price effects can generate spillovers because an older sibling's enrollment in college can increase the amount of financial aid a younger sibling is eligible for, as formulas for both federal financial aid and some colleges' own financial aid plans take such factors into account. Some colleges also offer same-campus tuition discounts for siblings. Importantly, price effects are salient only when both siblings are in college at the same point in time. When we split the sample by siblings' age difference, as in the first two columns of Table 6, we find that siblings one to three years apart exhibit somewhat smaller spillover effects than do siblings four or more years apart, even though the latter are unlikely to be enrolled in college simultaneously. Price effects therefore seem unlikely to explain a large portion of the observed spillovers.

Proximity effects occur because younger siblings may gain or lose utility from being on the same campus with their older siblings. This may stem from enjoying (or disliking) each other's company, providing study help to each other, providing information about campus-specific issues such as course choices, or even reducing enrollment costs through carpooling. That siblings four or more years apart exhibit large spillovers again make this channel less compelling, given that such siblings are unlikely to attend college simultaneously. That same-gender sibling pairs exhibit no larger spillovers than opposite-gender sibling pairs, as in the middle two columns of Table 6, also renders proximity effects less likely, assuming that former tend to be socially closer than the latter. Proximity effects also fail to explain why younger siblings' four-year college enrollment rate rose by twice as much as their target college enrollment rate, implying that many were induced to enroll in different four-year colleges than their older siblings. Finally, the last two columns of Table 6 show that spillover effects on four-year college enrollment and college quality are driven entirely by the subset of younger siblings who never even applied to their older siblings' target colleges.[18] Such sibling pairs could not attend the same campuses, making proximity effects an unlikely explanation for those spillovers.

Legacy effects occur because some colleges give admissions preferences to students whose family members have previously enrolled, in large part to increase the likelihood that family members donate money to the college (Hurwitz, 2011). Legacy effects are unlikely to explain the spillovers observed here, for two main reasons. First, the target colleges identified here are largely public, non-flagship institutions, whereas legacy admissions are concentrated in more elite private and public colleges. Second, as the discussion of proximity effects noted, much of the observed spillover effect runs through younger siblings attending different colleges than their older sibling and thus lacking any legacy status.

Having ruled out price, income, proximity and legacy effects as potential explanations for the observed college choice spillovers between siblings, we note that information effects are consistent with such patterns. An older sibling's enrollment at a particular college campus may generate information for parents or a younger sibling that would otherwise be costly or impossible to obtain. Older siblings attending higher quality colleges than they otherwise would may report to their families more positive experiences from those campuses. Higher persistence and graduation rates on such campuses may, in particular, change older siblings' reports of the returns to college enrollment. This is particularly true given that many older siblings would otherwise have attended two-year colleges with particularly low graduation rates.

Such information from older siblings may change younger siblings' own perceptions of the potential returns to attending high quality colleges, which would be consistent with our observations of younger siblings applying to and enrolling in higher quality colleges than they would otherwise. Younger siblings may place particularly high weight on their older siblings' college experiences, given that the educational success of a close family member may be more salient and predictive of one's own success than less personalized sources of information. Families may not actively seek out information about college quality and the expected returns from a given insti-

[18]Splitting the sample by an endogenous variable makes coefficient interpretation harder but provides one simple way to demonstrate spillovers between siblings not sharing a campus.

tution. If they do, the average statistics contained in websites and college guides may feel only somewhat relevant. Intimate personal knowledge of one person's positive experience with higher education may be substantially more powerful, particularly for those whose family characteristics suggest fewer close social connections to others pursuing higher education.

4 Conclusion

We provide some of the first causal evidence of peer effects in college choice, showing that younger siblings are more likely to enroll in college and choose higher quality colleges after their older siblings do so first. We argue that an older sibling's enrollment in a higher quality college can provide for families information about postsecondary education that would otherwise be difficult or impossible to obtain. In that sense, an older sibling's college choice is a particularly high touch intervention, providing prolonged exposure to another person's experience of the complex good that college education represents. Such information is particularly important in families where parents do not otherwise provide it, perhaps because they themselves have little or no college experience.

If students' college choices are deeply affected by the college experiences of people in their social networks, such social factors may partly explain persistent differences in college enrollment by income, race and geography. Further research might investigate the role played by college choices of other socially close individuals, such as parents, friends or neighbors. Such social factors have been relatively understudied by economists and might shed further light on the origins of inequalities in postsecondary outcomes.

References

Aguirre, J. and J. Matta (2018). Walking in your footsteps: Sibling spillovers in higher education choices. Working paper.

Altmejd, A., A. Barrios Fernandez, M. Drjle, C. Neilson, and D. Kovac (2019). Siblings' effects on university and major choice: Evidence from Chile, Croatia and Sweden. Working paper.

Altonji, J. G., S. Cattan, and I. Ware (2017). Identifying sibling influence on teenage substance use. *Journal of Human Resources 52*(1), 1–47.

Andrews, R. J., S. A. Imberman, and M. F. Lovenheim (2017). Risky business? The effect of majoring in business on earnings and educational attainment. Working Paper 23575, National Bureau of Economic Research.

Angrist, J. D. (2014). The perils of peer effects. *Labour Economics 30*, 98–108.

Barr, A. and B. Castleman (2018). An engine of economic opportunity: Intensive advising, college success, and social mobility. Working paper.

Belley, P. and L. Lochner (2007). The changing role of family income and ability in determining educational achievement. *Journal of Human Capital 1*(1), 37–89.

Bettinger, E. P., B. T. Long, P. Oreopoulos, and L. Sanbonmatsu (2012). The role of application assistance and information in college decisions: Results from the H&R Block FAFSA experiment. *The Quarterly Journal of Economics 127*(3), 1205–1242.

Bingley, P., P. Lundborg, and S. V. Lyk-Jensen (2019). Brothers in arms: Spillovers from a draft lottery. *Journal of Human Resources*, 0317–8646R3.

Bird, K. A., B. L. Castleman, J. T. Denning, J. Goodman, C. Lamberton, and K. O. Rosinger (2019). Nudging at scale: Experimental evidence from FAFSA completion campaigns. Working Paper 26158, National Bureau of Economic Research.

Black, S. E., S. Breining, D. N. Figlio, J. Guryan, K. Karbownik, H. S. Nielsen, J. Roth, and M. Simonsen (2017). Sibling spillovers. Working Paper 23062, National Bureau of Economic Research.

Black, S. E., K. E. Cortes, and J. A. Lincove (2015). Apply yourself: Racial and ethnic differences in college application. *Education Finance and Policy*, 1–59.

Campbell, S., L. Macmillan, R. Murphy, and G. Wyness (2019). Inequalities in student to course match: Evidence from linked administrative data. CEP Discussion Papers 1647.

Carrell, S. and B. Sacerdote (2017). Why do college-going interventions work? *American Economic Journal: Applied Economics 9*(3), 124–51.

Castleman, B. and J. Goodman (2018). Intensive college counseling and the enrollment and persistence of low-income students. *Education Finance and Policy 13*(1), 19–41.

Chetty, R., J. N. Friedman, and J. E. Rockoff (2014). Measuring the impacts of teachers II: Teacher value-added and student outcomes in adulthood. *American Economic Review 104*(9), 2633–79.

Chetty, R., J. N. Friedman, E. Saez, N. Turner, and D. Yagan (2017). Mobility report cards: The role of colleges in intergenerational mobility. Working Paper 23618, National Bureau of Economic Research.

Chetty, R., N. Hendren, P. Kline, and E. Saez (2014). Where is the land of opportunity? The geography of intergenerational mobility in the United States. *The Quarterly Journal of Economics 129*(4), 1553–1623.

Currie, J. and E. Moretti (2003). Mother's education and the intergenerational transmission of human capital: Evidence from college openings. *The Quarterly Journal of Economics 118*(4), 1495–1532.

Dahl, G. B., K. V. Løken, and M. Mogstad (2014). Peer effects in program participation. *American Economic Review 104*(7), 2049–74.

Deming, D. J., J. S. Hastings, T. J. Kane, and D. O. Staiger (2014). School choice, school quality, and postsecondary attainment. *American Economic Review 104*(3), 991–1013.

Dillon, E. W. and J. A. Smith (2017). Determinants of the match between student ability and college quality. *Journal of Labor Economics 35*(1), 45–66.

Dustan, A. (2018). Family networks and school choice. *Journal of Development Economics 134*, 372–391.

Goodman, J., O. Gurantz, and J. Smith (2020). Take two! SAT retaking and college enrollment gaps. *American Economic Journal: Economic Policy*.

Goodman, J., M. Hurwitz, and J. Smith (2017). Access to 4-year public colleges and degree completion. *Journal of Labor Economics 35*(3), 829–867.

Goodman, J., M. Hurwitz, J. Smith, and J. Fox (2015). The relationship between siblings' college choices: Evidence from one million SAT-taking families. *Economics of Education Review 48*, 75–85.

Gurantz, O., J. Howell, M. Hurwitz, C. Larson, M. Pender, and B. White (2019). Realizing your college potential? Impacts of College Board's RYCP campaign on postsecondary enrollment. Technical Report 40, Annenberg Institute at Brown University.

Hillman, N. W. (2016). Geography of college opportunity: The case of education deserts. *American Educational Research Journal 53*(4), 987–1021.

Hoxby, C. and C. Avery (2013). The missing "one-offs": The hidden supply of high-achieving, low-income students. *Brookings Papers on Economic Activity 44*(1), 1–65.

Hurwitz, M. (2011). The impact of legacy status on undergraduate admissions at elite colleges and universities. *Economics of Education Review 30*(3), 480–492.

Hurwitz, M. and J. Smith (2018). Student responsiveness to earnings data in the College Scorecard. *Economic Inquiry 56*(2), 1220–1243.

Hyman, J. (2019). Can light-touch college-going interventions make a difference? Evidence from a statewide experiment in Michigan. *Journal of Policy Analysis and Management*.

Imbens, G. and K. Kalyanaraman (2012). Optimal bandwidth choice for the regression discontinuity estimator. *The Review of Economic Studies 79*(3), 933–959.

Joensen, J. S. and H. S. Nielsen (2018). Spillovers in education choice. *Journal of Public Economics 157*, 158–183.

Lochner, L. and A. Monge-Naranjo (2012). Credit constraints in education. *Annual Review of Economics 4*(1), 225–256.

Loury, L. D. (2004). Siblings and gender differences in African-American college attendance. *Economics of Education Review 23*(3), 213–219.

Manski, C. F. (1993). Identification of endogenous social effects: The reflection problem. *The review of economic studies 60*(3), 531–542.

Mulhern, C. (2019a). Beyond teachers: Estimating individual guidance counselors' effects on educational attainment. Working paper.

Mulhern, C. (2019b). Changing college choices with personalized information at scale: Evidence on Naviance. Working paper.

Mwangi, C. A. G. (2015). (Re) examining the role of family and community in college access and choice: A metasynthesis. *The Review of Higher Education 39*(1), 123–151.

Oymak, C. (2018). High school students' views on who influences their thinking about education and careers. *National Center for Education Statistics, Stats in Brief. NCES 2018-088*.

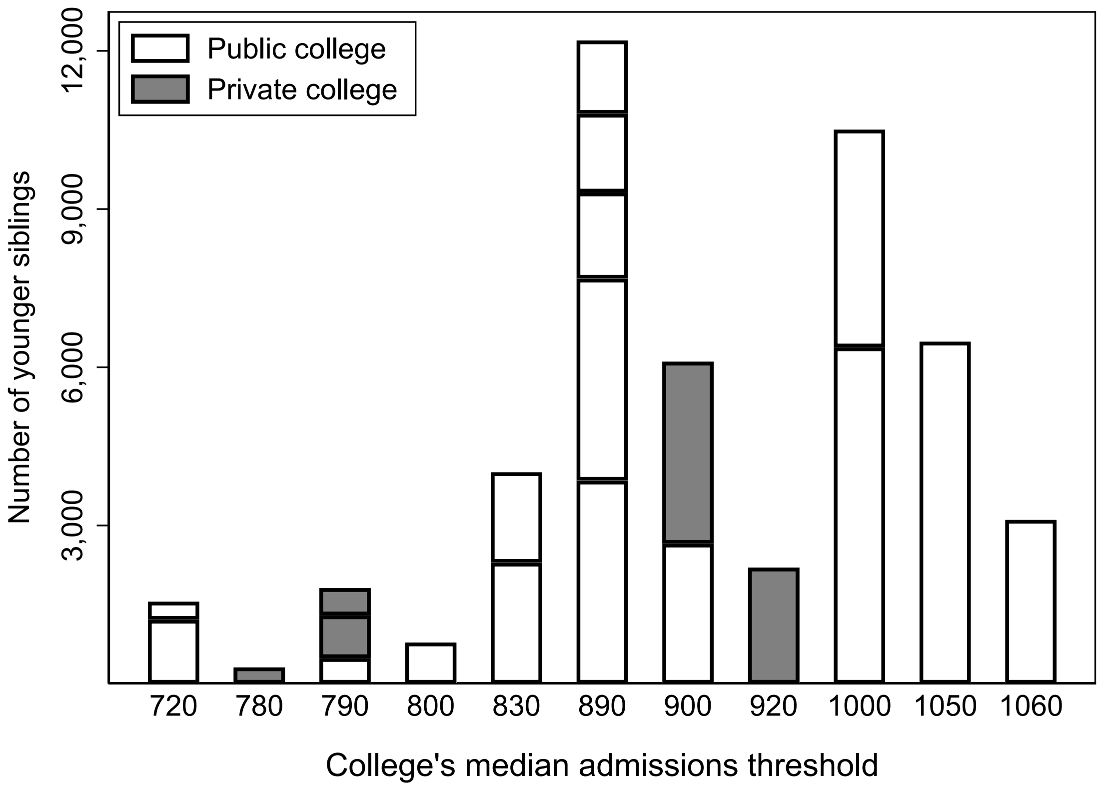
Qureshi, J. A. (2018). Additional returns to investing in girls' education: Impact on younger sibling human capital. *The Economic Journal 128*(616), 3285–3319.

Sacerdote, B. (2011). Peer effects in education: How might they work, how big are they and how much do we know thus far? In *Handbook of the Economics of Education*, Volume 3, pp. 249–277. Elsevier.

Smith, J. (2018). The sequential college application process. *Education Finance and Policy 13*(4), 545–575.

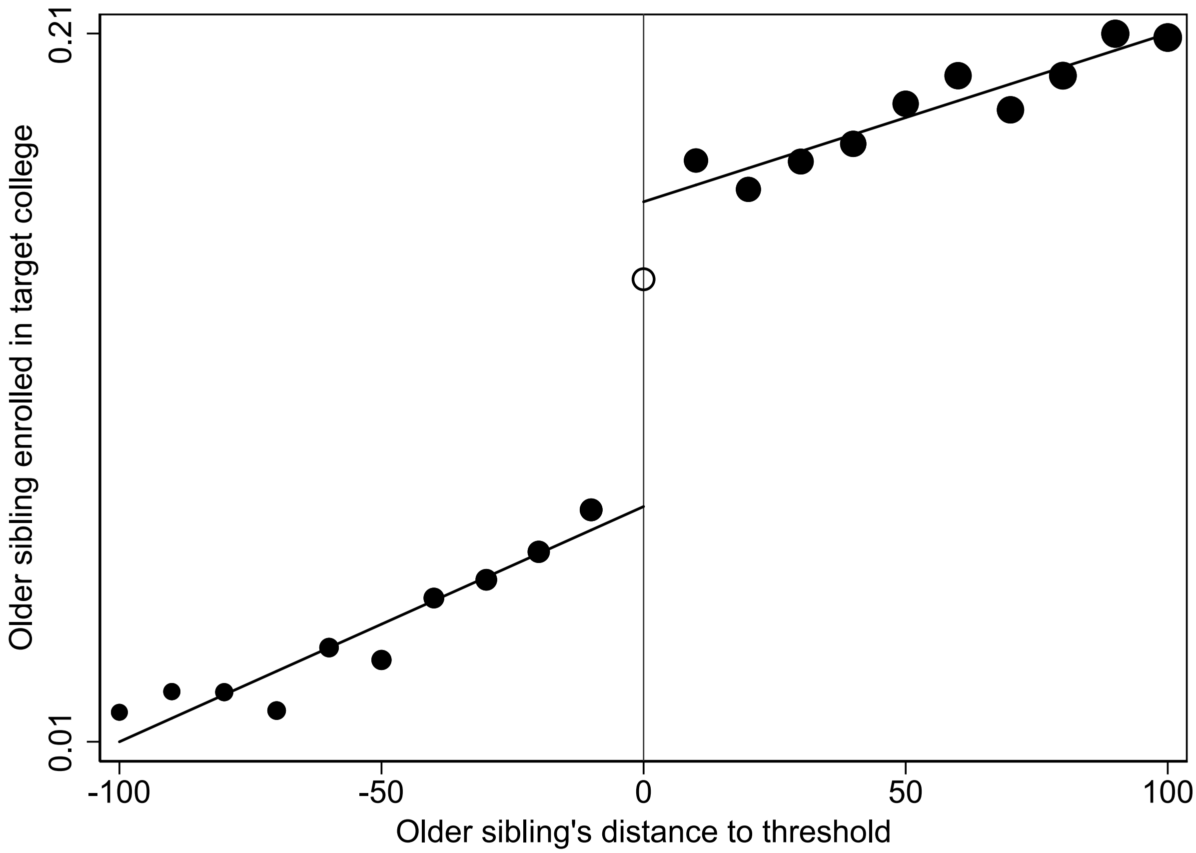
Smith, J. and K. Stange (2016). A new measure of college quality to study the effects of college sector and peers on degree attainment. *Education Finance and Policy 11*(4), 369–403.

Figure 1: Distribution of Younger Siblings Across Target Colleges



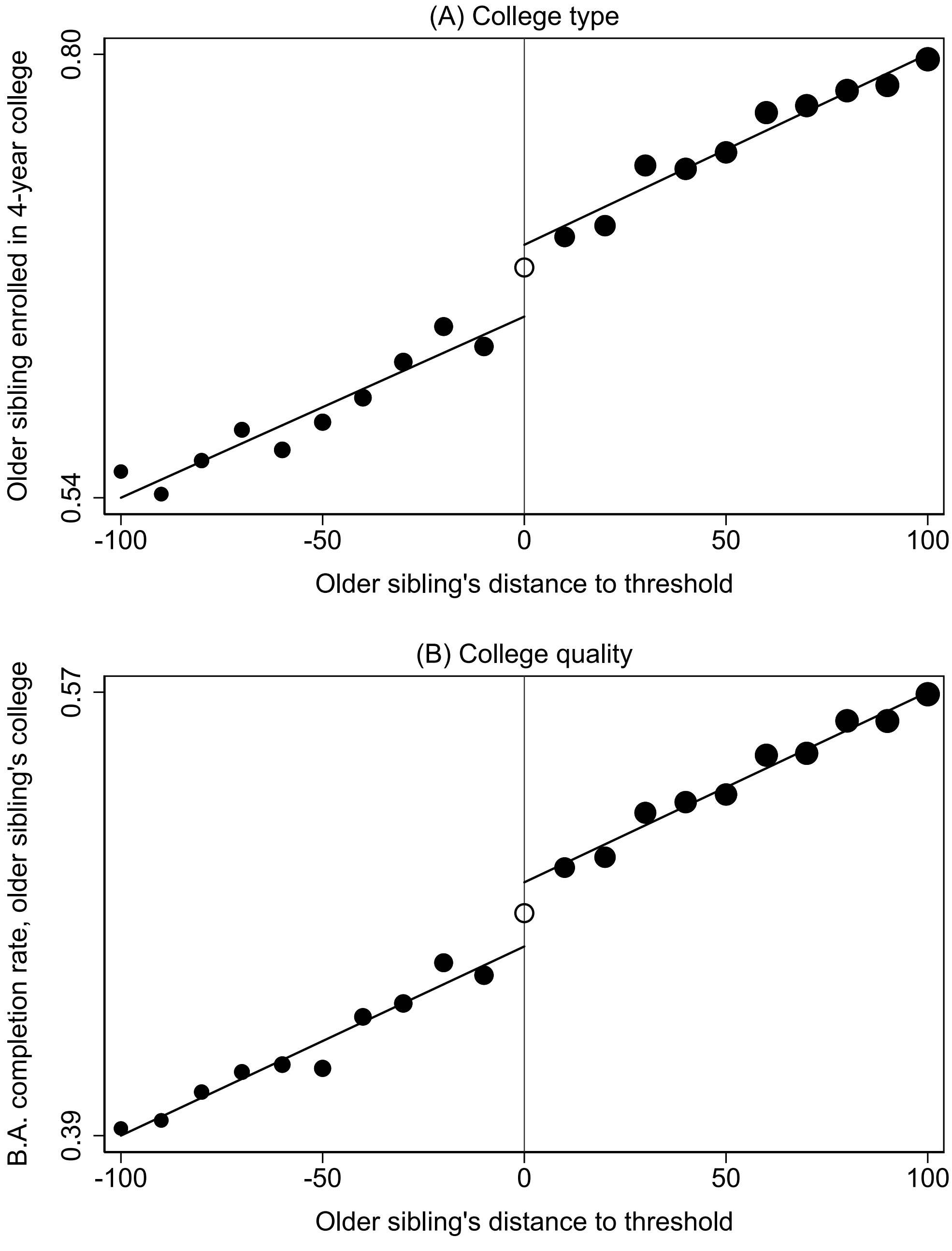
Notes: Shown above is the number of younger siblings in the main regression discontinuity sample, grouped by each college's median admissions threshold over the years observed. Each rectangle represents a single college.

Figure 2: Older Siblings' Enrollment in Target College



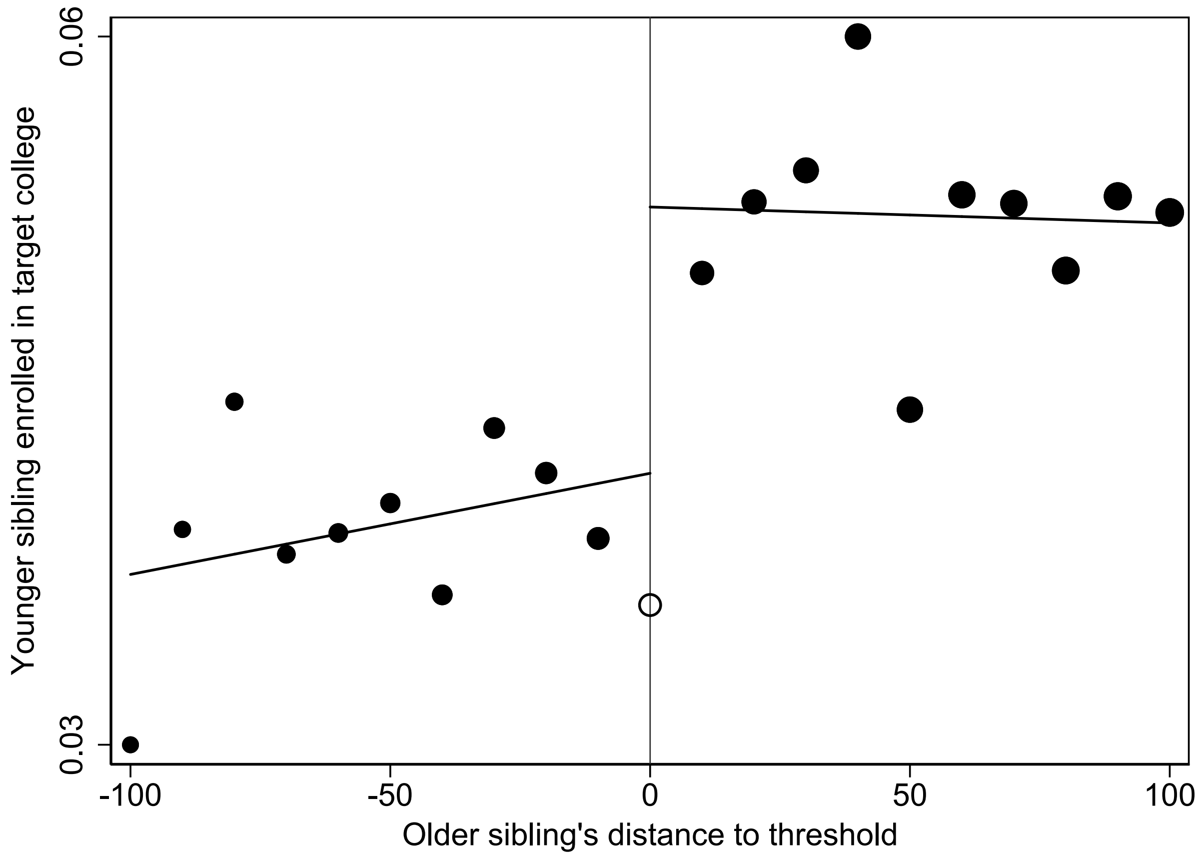
Notes: Shown above is older siblings' average rate of enrollment in the target college, by their own distance to the admissions threshold. Dot sizes are proportional to the number of younger siblings in each bin. Also shown are fitted lines from a local linear regression model using a bandwidth of 100 and excluding observations on the threshold.

Figure 3: Older Siblings' College Choices



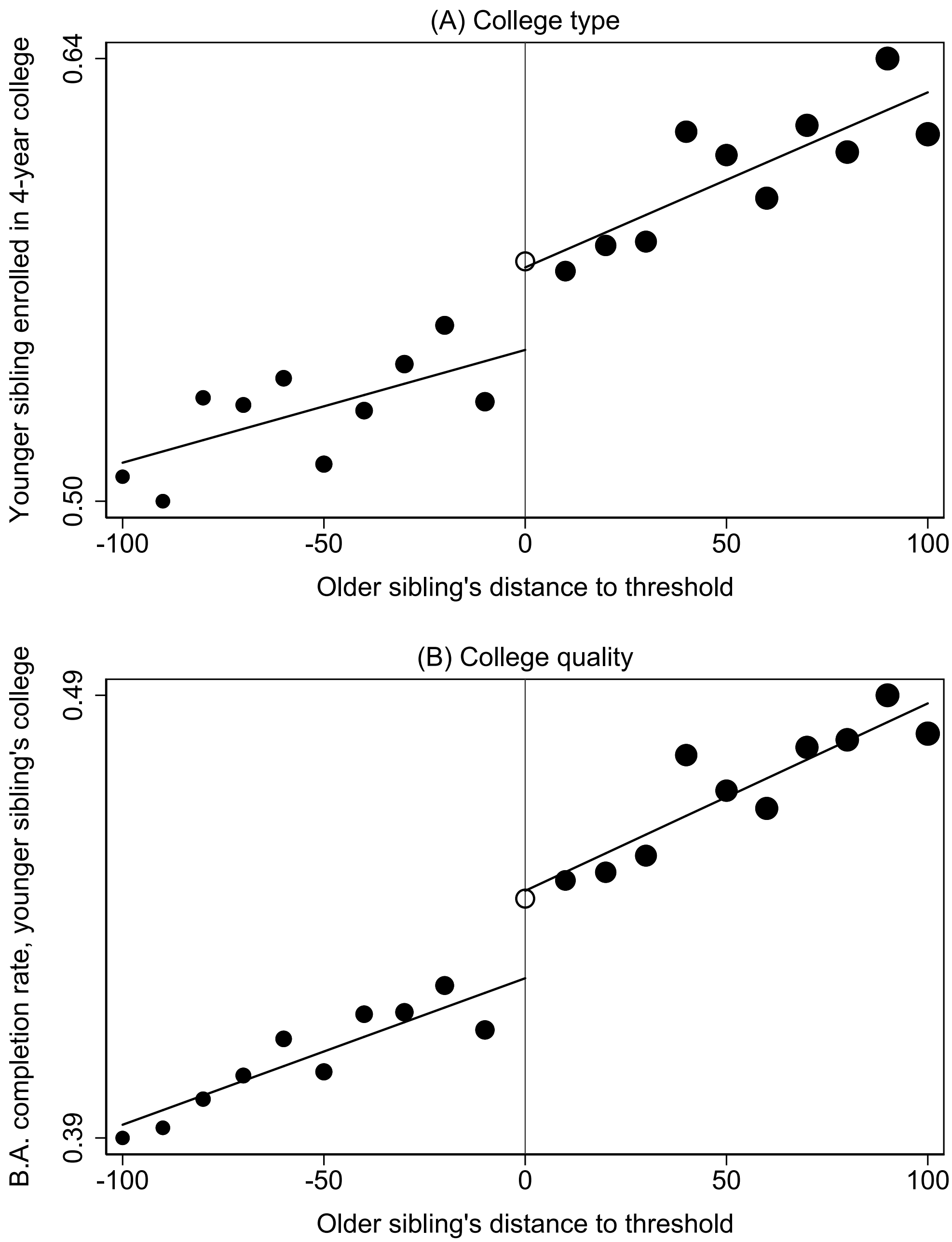
Notes: Shown above are older siblings' average enrollment rate in four-year college (panel A) and college quality measured by six-year B.A. completion rates (panel B), by their own distance to the admissions threshold. Dot sizes are proportional to the number of younger siblings in each bin. Also shown are fitted lines from a local linear regression model using a bandwidth of 100 and excluding observations on the threshold.

Figure 4: Younger Siblings' Enrollment in Older Siblings' Target Colleges



Notes: Shown above is younger siblings' average rate of enrollment in their older siblings' target colleges, by their older siblings' distance to the admissions threshold. Dot sizes are proportional to the number of younger siblings in each bin. Also shown are fitted lines from a local linear regression model using a bandwidth of 100 and excluding observations on the threshold.

Figure 5: Younger Siblings’ College Choices



Notes: Shown above are younger siblings’ average enrollment rate in four-year college (panel A) and college quality measured by six-year B.A. completion rates (panel B), by their older siblings’ distance to the admissions threshold. Dot sizes are proportional to the number of younger siblings in each bin. Also shown are fitted lines from a local linear regression model using a bandwidth of 100 and excluding observations on the threshold.

Table 1: Sample Characteristics

	Only children (1)	Sibling sample (2)	Older sibling applied to target college (3)	Regression discontinuity sample: All students (4)	Regression discontinuity sample: Uncertain college-goers (5)	Regression discontinuity sample: Probable college-goers (6)
(A) Demographics						
Siblings	1.00	2.32	2.26	2.25	2.25	2.25
White	0.51	0.66	0.66	0.58	0.38	0.67
Black	0.14	0.08	0.16	0.24	0.29	0.21
Hispanic	0.15	0.12	0.06	0.07	0.14	0.04
Asian	0.11	0.09	0.08	0.07	0.09	0.05
Income (000s)	67.1	80.0	82.5	75.4	43.8	92.1
Mother attended college	0.58	0.65	0.70	0.65	0.34	0.81
(B) Older siblings						
Maximum SAT score	1013	1082	1070	949	910	968
Score sends	3.67	4.37	6.61	6.15	5.61	6.41
Enrolled in 4-year college	0.54	0.68	0.78	0.70	0.60	0.75
College's B.A. completion rate	0.42	0.53	0.58	0.50	0.42	0.54
College's peer quality	0.06	0.25	0.26	0.05	-0.08	0.12
(C) Younger siblings						
Took SAT		0.83	0.88	0.85	0.76	0.90
Maximum SAT score		1072	1059	989	935	1012
Score sends		3.49	4.22	3.74	2.78	4.21
Enrolled in 4-year college		0.62	0.65	0.58	0.45	0.65
College's B.A. completion rate		0.49	0.51	0.45	0.35	0.50
College's peer quality		0.22	0.20	0.07	-0.12	0.16
N	11,287,116	3,145,006	166,080	49,069	16,356	32,713

Notes: Notes: Mean values of key variables are shown. Columns 1 and 2 divide the high school classes of 2004-14 into those with no observed siblings and those with at least one observed sibling. Column 3 includes only those families in which the oldest sibling applied to one of the target colleges. Column 4 limits that sample to those within 100 SAT points of the relevant threshold, excluding those on the threshold itself. Columns 5 and 6 divide the RD sample into those in the bottom third and top two-thirds of the distribution of predicted four-year college enrollment. College quality is measured by the fraction of students starting at that college who complete a B.A. anywhere within six years and the mean standardized PSAT score of students at that college.

Table 2: First Stage Impacts of Thresholds on Older Siblings' Target College Enrollment

Bandwidth =	60	80	100	120	140	100	100
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
All students	0.075***	0.082***	0.085***	0.095***	0.101***	0.085***	0.061***
	(0.009)	(0.008)	(0.007)	(0.006)	(0.006)	(0.007)	(0.007)
F-statistic	62.4	109.3	152.9	234.8	314.1	154.0	80.8
Uncertain college-goers	0.078***	0.076***	0.083***	0.091***	0.101***	0.083***	0.060***
	(0.016)	(0.013)	(0.012)	(0.010)	(0.010)	(0.012)	(0.012)
F-statistic	23.2	33.1	50.6	76.0	113.0	50.9	26.5
Probable college-goers	0.073***	0.084***	0.085***	0.095***	0.100***	0.085***	0.061***
	(0.011)	(0.010)	(0.008)	(0.008)	(0.007)	(0.008)	(0.008)
F-statistic	40.2	76.7	100.0	160.7	205.8	101.2	54.8
Donut hole	Y	Y	Y	Y	Y	Y	N
Demographic controls	N	N	N	N	N	Y	N

Notes: Heteroskedasticity robust standard errors clustered by oldest sibling's high school are in parentheses (* p<.10 ** p<.05 *** p<.01). Each coefficient is a first stage estimate of the impact of an older sibling's threshold-based admissibility on their enrollment in the target college. Each estimate comes from a local linear regression that includes fixed effects for each combination of older sibling's cohort, younger sibling's cohort, and the target college to which the older sibling applied. Columns 1-5 use bandwidths from 60 to 140 SAT points and a donut hole specification that exclude observations on the threshold itself. Columns 6 and 7 replicate column 3 but respectively add demographic controls (for gender, race, income and parental education) and eliminate the donut hole. The first row includes all students, while the second and third rows divide the sample into those in the bottom third and top two-thirds of the distribution of predicted four-year college enrollment. Beneath each coefficient is the F-statistic for the admissibility instrument.

Table 3: Characteristics of Older Siblings' Chosen Colleges

	College type		College quality		Price, location	
	4-year college (1)	2-year college (2)	B.A. completion rate (3)	Peer quality (Z-score) (4)	Net price (000s) (5)	50+ miles from home (6)
All students	0.424***	-0.344***	0.257***	0.446***	3.289	0.324***
	(0.115)	(0.105)	(0.063)	(0.103)	(2.026)	(0.125)
Control complier $\hat{Y}$	0.58	0.34	0.41	-0.06	10.44	0.23
Uncertain college-goers	0.409**	-0.465**	0.317***	0.570***	1.920	0.336
	(0.207)	(0.196)	(0.110)	(0.185)	(3.220)	(0.207)
Control complier $\hat{Y}$	0.59	0.47	0.31	-0.29	11.37	0.21
Probable college-goers	0.370***	-0.229*	0.197**	0.361***	3.356	0.289*
	(0.140)	(0.126)	(0.078)	(0.127)	(2.642)	(0.157)
Control complier $\hat{Y}$	0.63	0.23	0.49	0.07	10.57	0.28

Notes: Heteroskedasticity robust standard errors clustered by oldest sibling's high school are in parentheses (* p<.10 ** p<.05 *** p<.01). Each coefficient is an instrumental variables estimate of the impact of an older sibling's enrollment in the target college on their own college choices, using admissibility as an instrument. Each estimate comes from a local linear regression with a bandwidth of 100 SAT points, a donut hole specification that excludes observations on the threshold, and fixed effects for each combination of older sibling's cohort, younger sibling's cohort, and older sibling's target college. The first row includes all students, while the second and third rows divide the sample into those in the bottom third and top two-thirds of the distribution of predicted four-year college enrollment. College quality is measured by the fraction of students starting at that college who complete a B.A. anywhere within six years (column 3) and the mean standardized PSAT score of students at that college (column 4). Also listed below each coefficient is the predicted value of the outcome for control compliers.

Table 4: Younger Siblings' College Entrance Exam and Application Choices

	College entrance exams		College applications		
	Ever took SAT (1)	Maximum SAT score (2)	Target college (3)	Total number (4)	Only high grad. rate colleges (5)
All students	0.074	41.1	0.267***	0.610	0.278**
	(0.090)	(39.8)	(0.096)	(0.910)	(0.119)
Control complier $\hat{Y}$	0.82	949	0.10	2.75	0.25
Uncertain college-goers	0.001	81.7	0.250	0.248	0.398*
	(0.178)	(68.7)	(0.164)	(1.379)	(0.205)
Control complier $\hat{Y}$	0.87	858	0.13	2.97	0.10
Probable college-goers	0.113	-0.5	0.285**	0.583	0.196
	(0.095)	(50.7)	(0.124)	(1.197)	(0.152)
Control complier $\hat{Y}$	0.80	1015	0.06	2.78	0.35

Notes: Heteroskedasticity robust standard errors clustered by oldest sibling's high school are in parentheses (* $p<.10$ ** $p<.05$ *** $p<.01$). Each coefficient is an instrumental variables estimate of the impact of an older sibling's enrollment in the target college on younger siblings' testing and application choices, using admissibility as an instrument. Each estimate comes from a local linear regression that includes fixed effects for each combination of older sibling's cohort, younger sibling's cohort, and the target college to which the older sibling applied. Each regression uses a bandwidth of 100 SAT points and a donut hole specification that exclude observations on the threshold itself. The first row includes all students, while the second and third rows divide the sample into those in the bottom third and top two-thirds of the distribution of predicted four-year college enrollment. The outcome in column 5 is an indicator for a student's application portfolio containing only colleges whose B.A. completion rates exceed 50 percent. Also listed below each coefficient is the predicted value of the outcome for control compliers.

Table 5: Younger Siblings' College Choices

		College type		College quality		Price, location	
	Followed to target college	4-year college	2-year college	B.A. completion rate	Peer quality (Z-score)	Net price (000s)	50+ miles from home
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
All students	0.132***	0.275**	-0.060	0.201***	0.345**	3.556	0.190
	(0.048)	(0.122)	(0.101)	(0.074)	(0.137)	(2.191)	(0.117)
Control complier $\hat{Y}$	0.02	0.33	0.26	0.28	-0.24	7.51	0.19
Uncertain college-goers	0.200**	0.448**	0.061	0.423***	0.649***	9.820***	0.459**
	(0.085)	(0.216)	(0.186)	(0.129)	(0.230)	(3.614)	(0.198)
Control complier $\hat{Y}$	0.00	0.13	0.12	0.03	-0.65	0.86	0.00
Probable college-goers	0.096	0.152	-0.106	0.068	0.158	-0.399	0.010
	(0.062)	(0.150)	(0.128)	(0.091)	(0.170)	(2.824)	(0.150)
Control complier $\hat{Y}$	0.06	0.48	0.31	0.44	0.01	11.81	0.36

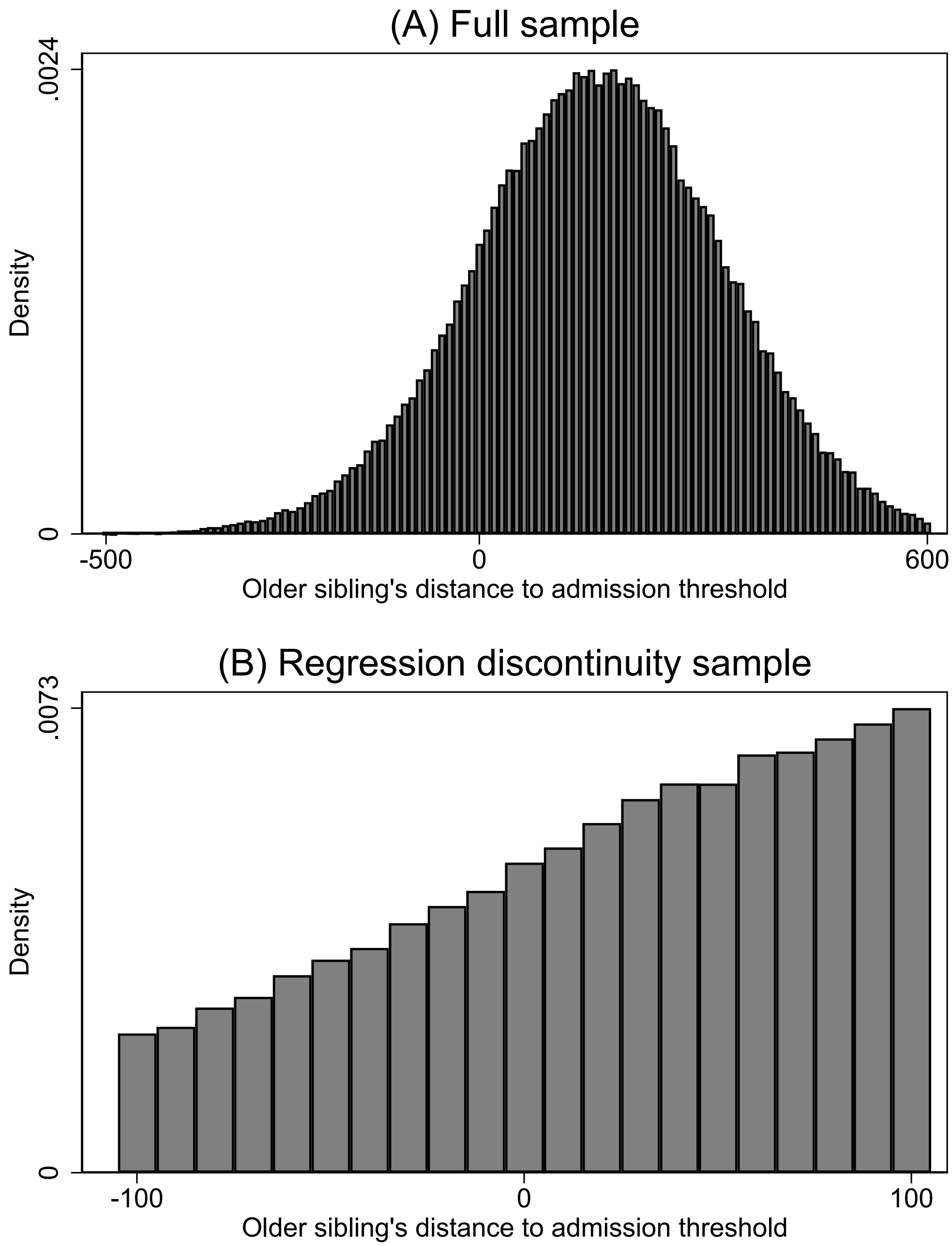
Notes: Heteroskedasticity robust standard errors clustered by oldest sibling's high school are in parentheses (* p<.10 ** p<.05 *** p<.01). Each coefficient is an instrumental variables estimate of the impact of an older sibling's enrollment in the target college on younger siblings' college choices, using admissibility as an instrument. Each estimate comes from a local linear regression with a bandwidth of 100 SAT points, a donut hole specification that excludes observations on the threshold, and fixed effects for each combination of older sibling's cohort, younger sibling's cohort, and older sibling's target college. The first row includes all students, while the second and third rows divide the sample into those in the bottom third and top two-thirds of the distribution of predicted four-year college enrollment. College quality is measured by the fraction of students starting at that college who complete a B.A. anywhere within six years (column 4) and the mean standardized PSAT score of students at that college (column 5). Also listed below each coefficient is the predicted value of the outcome for control compliers.

Table 6: Exploring Channels for Spillover Effects

	Siblings' age difference		Siblings' genders		Younger sibling applied to target college	
	1-3 years (1)	4+ years (2)	Same (3)	Different (4)	No (5)	Yes (6)
(A) All students						
Target college	0.108*	0.195**	0.072	0.176**	0.054*	0.145
	(0.057)	(0.096)	(0.065)	(0.070)	(0.031)	(0.133)
Four-year college	0.176	0.529**	0.217	0.326*	0.387**	-0.121
	(0.135)	(0.247)	(0.165)	(0.175)	(0.169)	(0.149)
College's B.A. completion rate	0.166**	0.290**	0.168*	0.226**	0.277***	-0.068
	(0.081)	(0.147)	(0.099)	(0.108)	(0.103)	(0.087)
N	33,356	15,713	25,047	24,016	39,142	9,865
(B) Uncertain college-goers						
Target college	0.179*	0.246	0.154	0.295**	0.097	0.337
	(0.099)	(0.165)	(0.112)	(0.137)	(0.064)	(0.225)
Four-year college	0.293	0.812*	0.328	0.674*	0.600*	-0.120
	(0.240)	(0.459)	(0.277)	(0.359)	(0.307)	(0.259)
College's B.A. completion rate	0.355**	0.579**	0.375**	0.498**	0.550***	0.011
	(0.142)	(0.276)	(0.165)	(0.218)	(0.185)	(0.147)
N	10,866	5,482	8,461	7,844	13,407	2,815

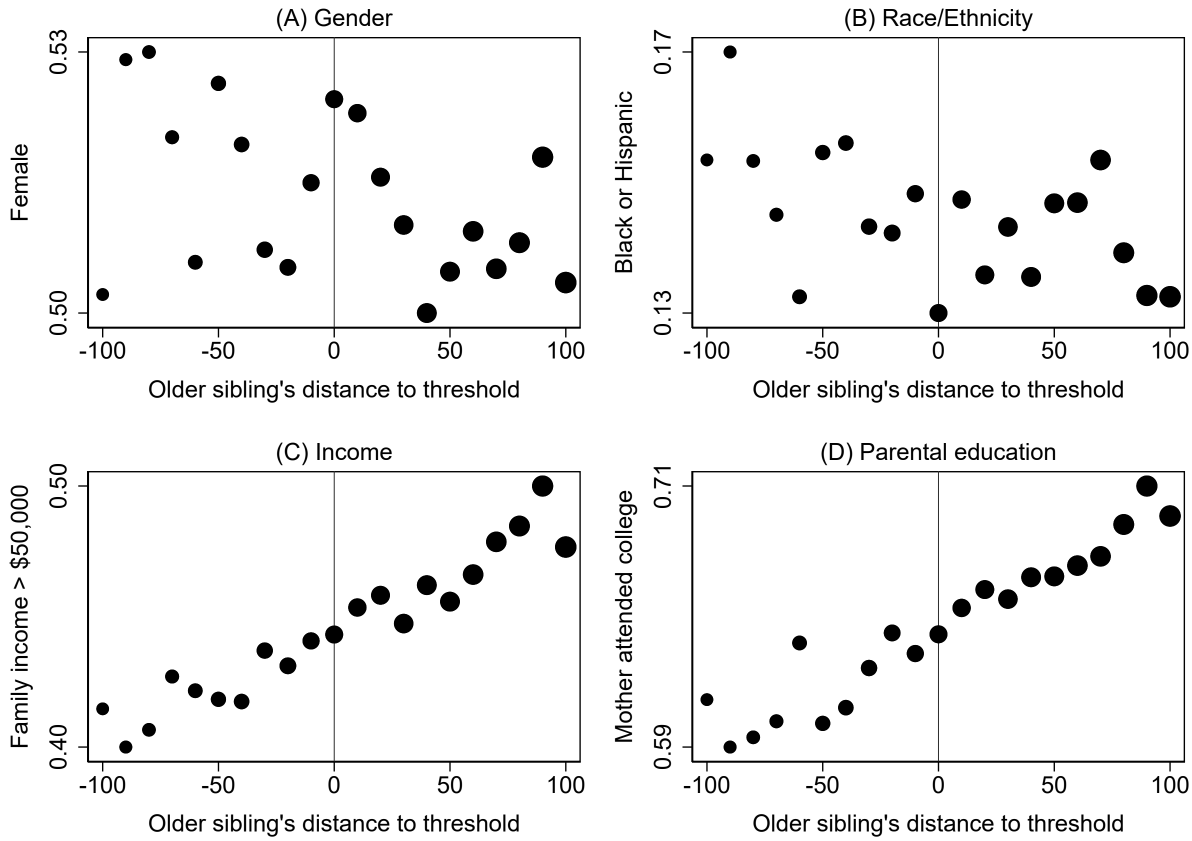
Notes: Heteroskedasticity robust standard errors clustered by oldest sibling's high school are in parentheses (* p<.10 ** p<.05 *** p<.01). Each coefficient is an instrumental variables estimate of the impact of an older sibling's enrollment in the target college on younger siblings' college choices, using admissibility as an instrument. Each estimate comes from a local linear regression with a bandwidth of 100 SAT points, a donut hole specification that excludes observations on the threshold, and fixed effects for each combination of older sibling's cohort, younger sibling's cohort, and older sibling's target college. Columns 1 and 2 split the sample by siblings' age difference, columns 3 and 4 by siblings' gender match, and columns 5 and 6 by whether the younger sibling applied to the older sibling's target college. Panel A includes all students while panel B includes only those whose family characteristics predict four-year college enrollment probabilities in the bottom third of the distribution.

Figure A.1: Density of Observations



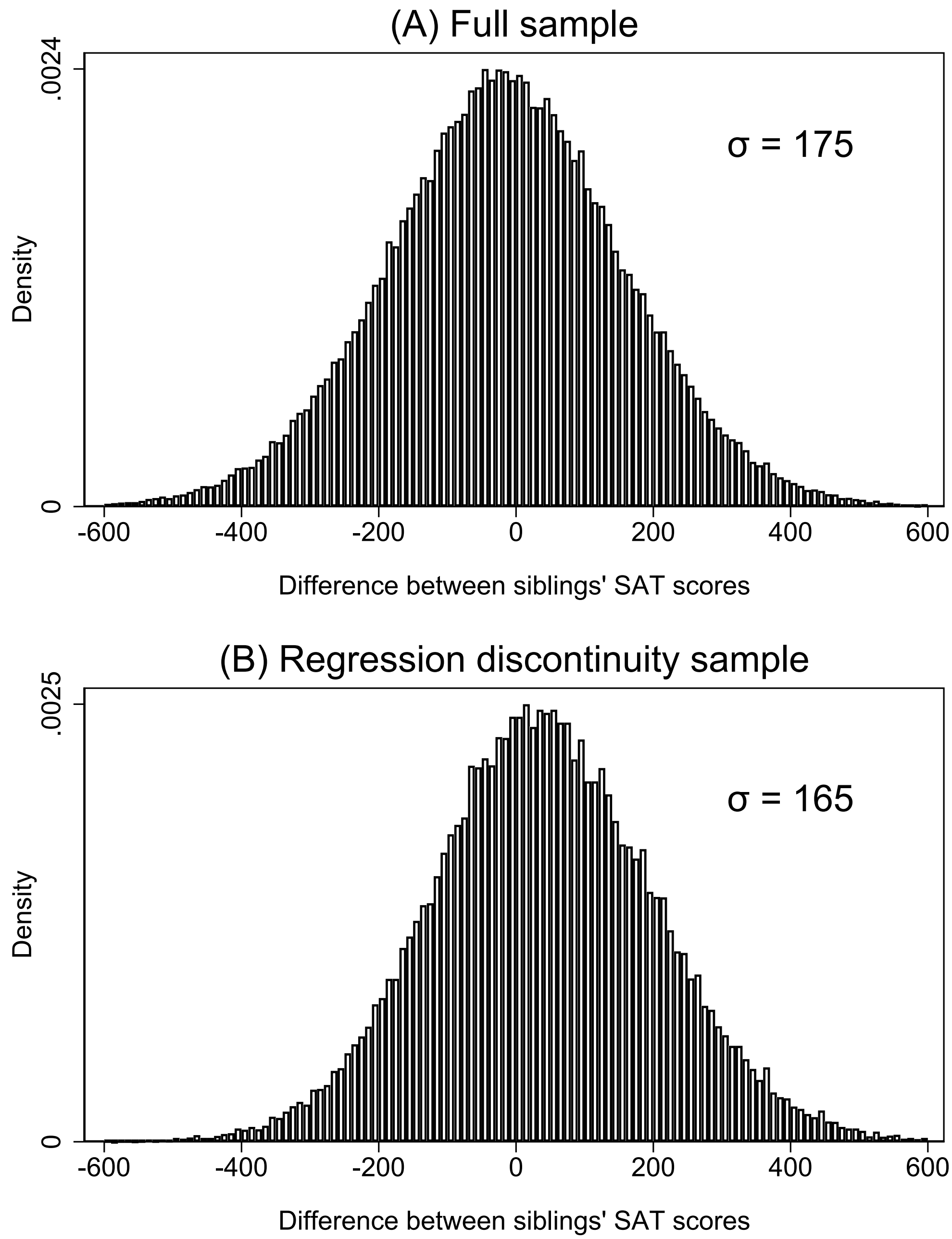
Notes: Shown above is the density of younger siblings by their older siblings' distance to the admissions threshold. Panel A shows the full sample of those whose older siblings applied to a target college, while panel B limits the sample to those whose older siblings' SAT scores were within 100 points of the relevant threshold.

Figure A.2: Covariate Balance



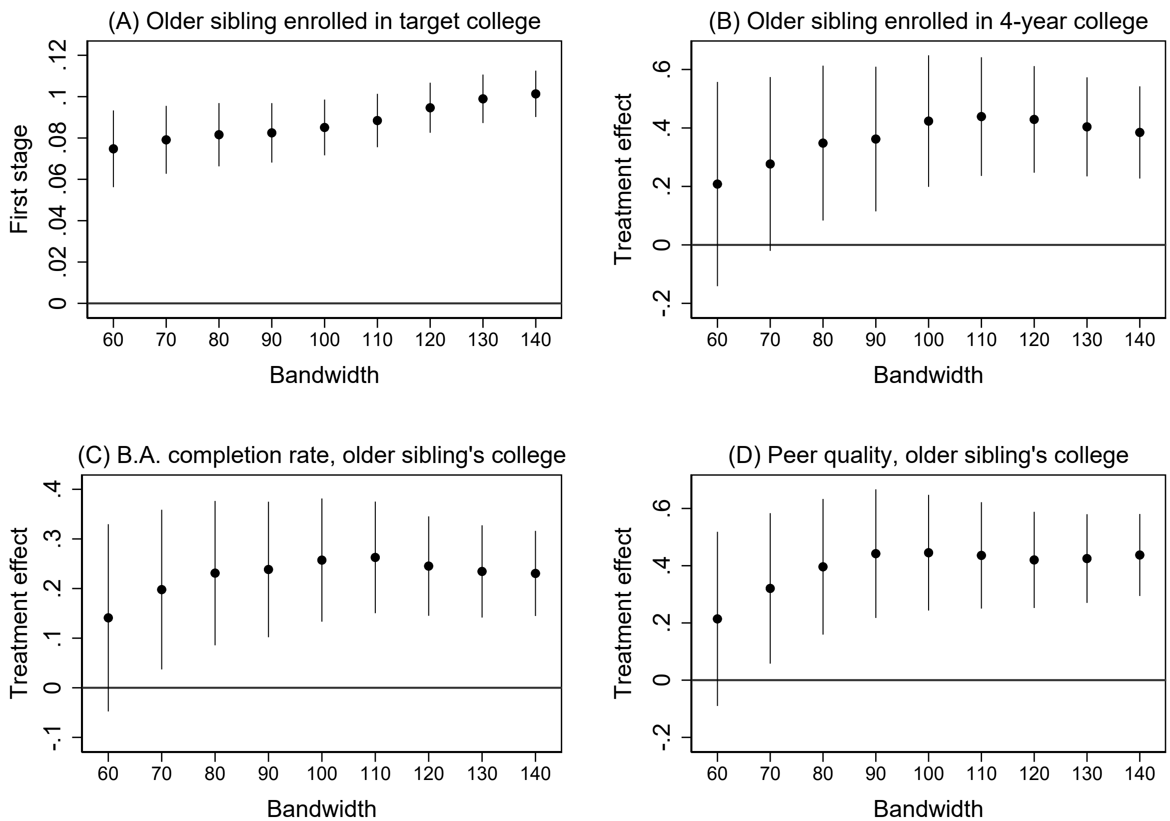
Notes: Shown above are the mean values of younger siblings' characteristics by their older siblings' distance to the admissions threshold. Dot sizes are proportional to the number of younger siblings in each bin.

Figure A.3: Differences in Siblings' SAT Scores



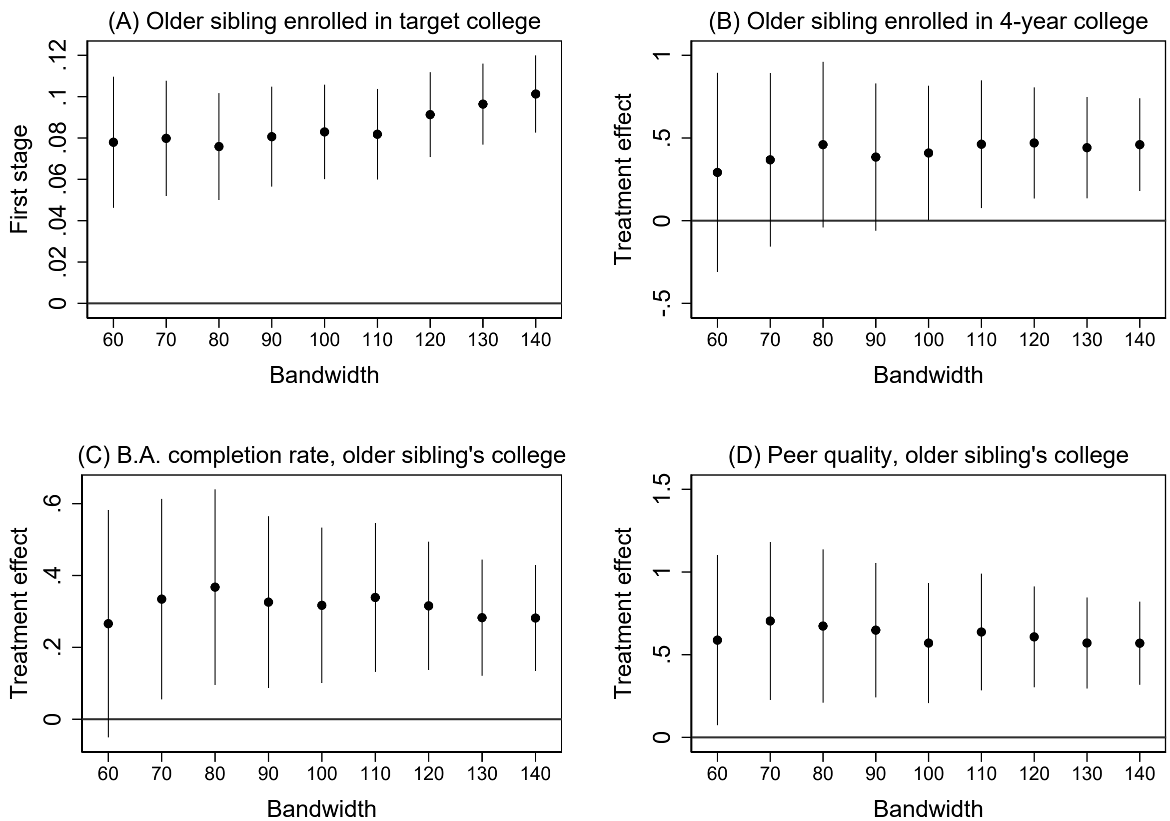
Notes: Shown above is the density of the difference between younger and older siblings' SAT scores (for younger siblings who took the SAT). Also shown is the standard deviation of that distribution. Panel A shows the full sample of those whose older siblings applied to a target college, while panel B limits the sample to those whose older siblings' SAT scores were within 100 points of the relevant threshold.

Figure A.4: Robustness of Older Sibling Effects to Bandwidth Choice, All Students



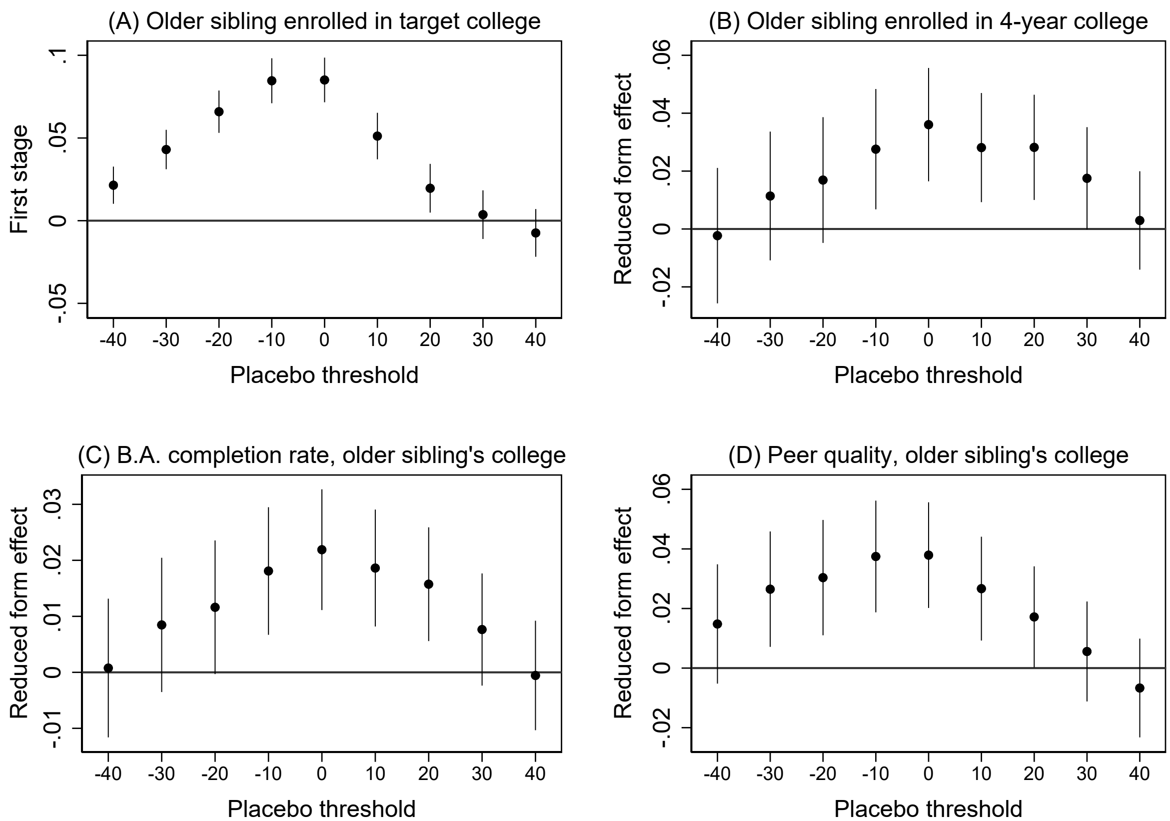
Notes: Shown above are first stage estimates of older siblings' admissibility on their target college enrollment (panel A) and instrumental variables estimates (panels B-D) of the impact of older siblings' target college enrollment on their own college choices. Also shown are 95 percent confidence intervals based on heteroskedasticity robust standard errors clustered by oldest sibling's high school. Each estimate comes from a local linear regression with the listed bandwidth, a donut hole specification that excludes observations on the threshold, and fixed effects for each combination of older sibling's cohort, younger sibling's cohort, and older sibling's target college. The sample includes all students.

Figure A.5: Robustness of Older Sibling Effects to Bandwidth Choice, Uncertain College-Goers



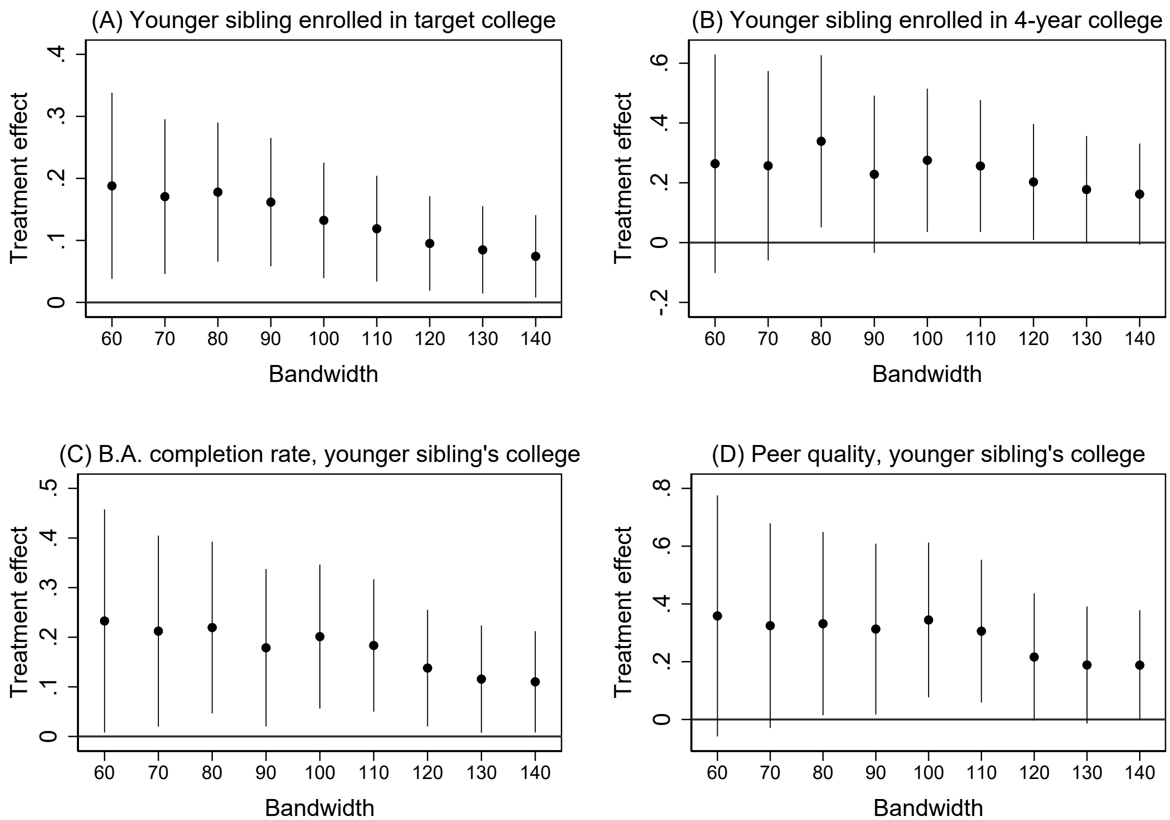
Notes: Shown above are instrumental variables estimates of the impact of older siblings' target college enrollment on their own college choices. Also shown are 95 percent confidence intervals based on heteroskedasticity robust standard errors clustered by oldest sibling's high school. Each estimate comes from a local linear regression with the listed bandwidth, a donut hole specification that excludes observations on the threshold, and fixed effects for each combination of older sibling's cohort, younger sibling's cohort, and older sibling's target college. The sample includes only students whose family characteristics predict four-year college enrollment probabilities in the bottom third of the distribution.

Figure A.6: Placebo Thresholds for Older Sibling Effects



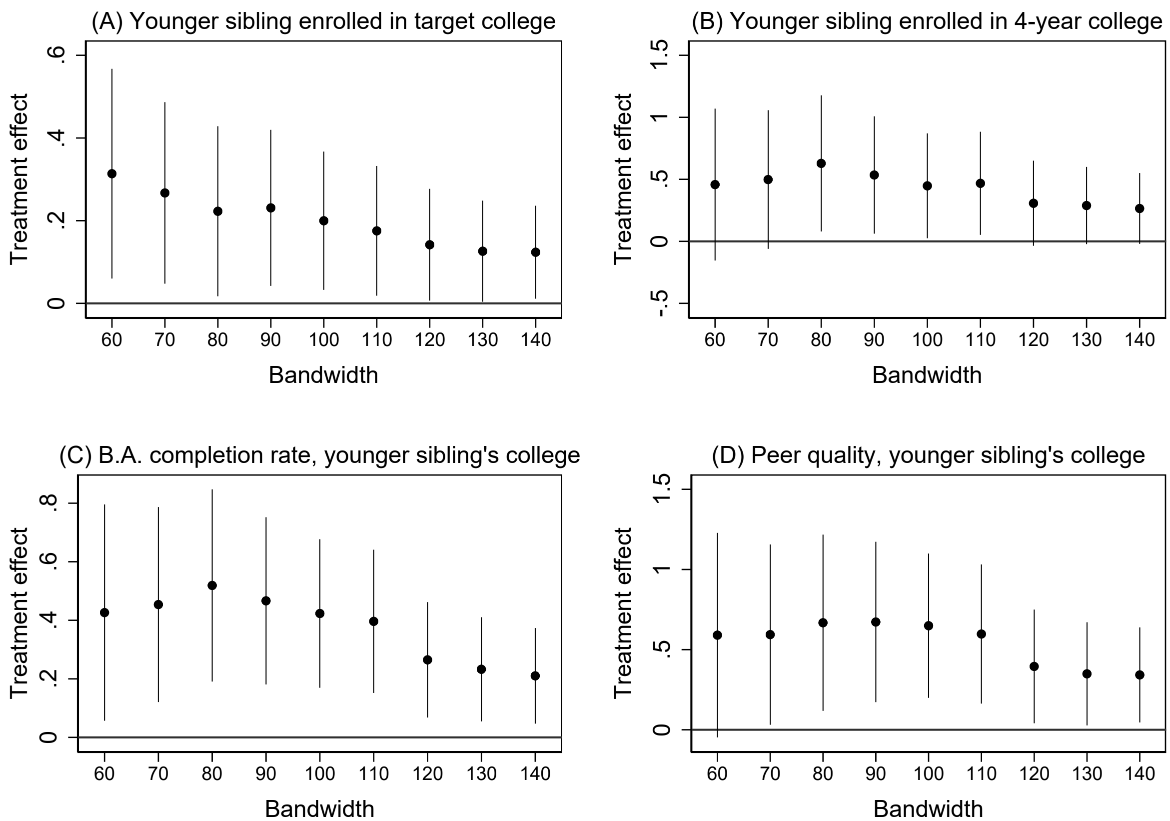
Notes: Shown above are first stage (panel A) and reduced form (panels B-D) estimates of older siblings' admissibility on their own college choices. Also shown are 95 percent confidence intervals based on heteroskedasticity robust standard errors clustered by oldest sibling's high school. Each estimate comes from a local linear regression re-centered around the true admissions threshold by the listed amount, with a bandwidth of 100 SAT points, a donut hole specification that excludes observations on the re-centered threshold, and fixed effects for each combination of older sibling's cohort, younger sibling's cohort, and older sibling's target college. The sample includes all students.

Figure A.7: Robustness of Younger Sibling Effects to Bandwidth Choice, All Students



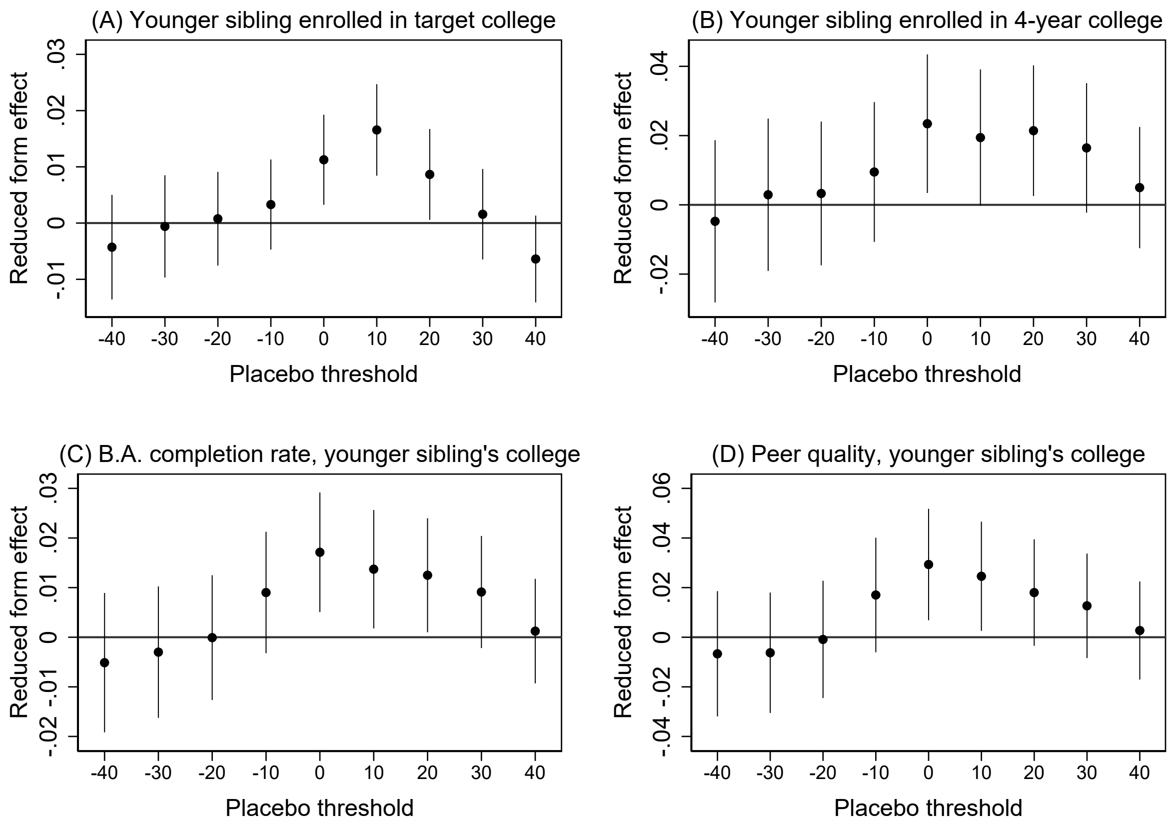
Notes: Shown above are instrumental variables estimates of the impact of older siblings' target college enrollment on the characteristics of younger siblings' chosen colleges. Also shown are 95 percent confidence intervals based on heteroskedasticity robust standard errors clustered by oldest sibling's high school. Each estimate comes from a local linear regression with the listed bandwidth, a donut hole specification that excludes observations on the threshold, and fixed effects for each combination of older sibling's cohort, younger sibling's cohort, and older sibling's target college. The sample includes all students.

Figure A.8: Robustness of Younger Sibling Effects to Bandwidth Choice, Uncertain College-Goers



Notes: Shown above are instrumental variables estimates of the impact of older siblings' target college enrollment on younger siblings' college choices. Also shown are 95 percent confidence intervals based on heteroskedasticity robust standard errors clustered by oldest sibling's high school. Each estimate comes from a local linear regression with the listed bandwidth, a donut hole specification that excludes observations on the threshold, and fixed effects for each combination of older sibling's cohort, younger sibling's cohort, and older sibling's target college. The sample includes only students whose family characteristics predict four-year college enrollment probabilities in the bottom third of the distribution.

Figure A.9: Placebo Thresholds for Younger Sibling Effects



Notes: Shown above are reduced form estimates of older siblings' admissibility on their younger siblings' college choices. Also shown are 95 percent confidence intervals based on heteroskedasticity robust standard errors clustered by oldest sibling's high school. Each estimate comes from a local linear regression re-centered around the true admissions threshold by the listed amount, with a bandwidth of 100 SAT points, a donut hole specification that excludes observations on the re-centered threshold, and fixed effects for each combination of older sibling's cohort, younger sibling's cohort, and older sibling's target college. The sample includes all students.

Table A.1: Covariate Balance Tests

	Female (1)	White (2)	Black or Hispanic (3)	Income above 30,000 (4)	Income above 50,000 (5)	Income above 75,000 (6)	Income response missing (7)	Mother attended college (8)	Predicted prob. of college (9)
All students	0.004	-0.006	-0.001	0.008	-0.001	0.000	-0.004	0.001	0.003
	(0.010)	(0.010)	(0.009)	(0.011)	(0.011)	(0.010)	(0.010)	(0.011)	(0.002)
Control mean	0.52	0.54	0.34	0.55	0.44	0.33	0.33	0.63	0.71
N	49,069	49,069	49,069	49,069	49,069	49,069	49,069	49,069	49,069
Uncertain college-goers	0.001	-0.013	0.002	0.005	0.000	-0.006	0.018	-0.032*	-0.001
	(0.017)	(0.017)	(0.017)	(0.019)	(0.016)	(0.012)	(0.017)	(0.017)	(0.002)
Control mean	0.53	0.36	0.45	0.43	0.24	0.11	0.29	0.36	0.61
N	16,348	16,348	16,348	16,348	16,348	16,348	16,348	16,348	16,348
Probable college-goers	0.005	-0.008	0.002	0.006	-0.006	-0.001	-0.019	0.010	0.002
	(0.013)	(0.011)	(0.010)	(0.013)	(0.013)	(0.013)	(0.013)	(0.010)	(0.001)
Control mean	0.50	0.65	0.28	0.62	0.56	0.47	0.35	0.80	0.76
N	32,686	32,686	32,686	32,686	32,686	32,686	32,686	32,686	32,686

Notes: Heteroskedasticity robust standard errors clustered by oldest sibling's high school are in parentheses (* $p<.10$ ** $p<.05$ *** $p<.01$). All coefficients are reduced form estimates of an older sibling's admissibility to the target college on the covariate listed. Each estimate comes from a local linear regression that includes fixed effects for each combination of older sibling's cohort, younger sibling's cohort, and the target college to which the older sibling applied. Each regression uses a bandwidth of 100 SAT points and excludes observations on the threshold itself. The outcome in column 9 is the younger sibling's predicted probability of enrolling in four-year college. Also listed are mean values of the covariate for students just below the threshold.

Table A.2: Further Robustness of Younger Siblings' College Choices

	Baseline specification (1)	Including covariates (2)	Eliminating donut hole (3)
(A) All students			
Enrolled in target college	0.132***	0.128***	0.223***
	(0.048)	(0.047)	(0.062)
Enrolled in 4-year college	0.275**	0.219*	0.205
	(0.122)	(0.116)	(0.150)
B.A. completion rate	0.201***	0.159**	0.175*
	(0.074)	(0.069)	(0.092)
Peer quality	0.345**	0.258**	0.339**
	(0.137)	(0.129)	(0.170)
(B) Uncertain college-goers			
Enrolled in target college	0.200**	0.200**	0.332***
	(0.085)	(0.085)	(0.118)
Enrolled in 4-year college	0.448**	0.454**	0.472*
	(0.216)	(0.213)	(0.272)
B.A. completion rate	0.423***	0.407***	0.472***
	(0.129)	(0.127)	(0.171)
Peer quality	0.649***	0.590***	0.784***
	(0.230)	(0.223)	(0.304)

Notes: Heteroskedasticity robust standard errors clustered by oldest sibling's high school are in parentheses (* p<.10 ** p<.05 *** p<.01). Each coefficient is an instrumental variables estimate of the impact of an older sibling's enrollment in the target college on younger siblings' college choices, using admissibility as an instrument. Each estimate comes from a local linear regression that includes fixed effects for each combination of older sibling's cohort, younger sibling's cohort, and the target college to which the older sibling applied. Column 1 uses a bandwidth of 100 SAT points and a donut hole specification that exclude observations on the threshold itself. Column 2 adds to that regression covariates, including gender, race, income and parental education. Column 3 eliminates the donut hole, including observations on the threshold itself. Panel A includes all students, while panel B includes those in the bottom third of the distribution of predicted four-year college enrollment.