The Relationship Between Siblings' College Choices: Evidence from One Million SAT-Taking Families

Faculty Research Working Paper Series

Joshua Goodman
Harvard Kennedy School

Michael Hurwitz
College Board

Jonathan Smith
College Board

Julia Fox
College Board

September 2014
RWP14-043

The Relationship Between Siblings' College Choices:
Evidence from One Million SAT-Taking Families*

Joshua Goodman
Harvard University
joshua_goodman@hks.harvard.edu

Michael Hurwitz
College Board
mhurwitz@collegeboard.org

Jonathan Smith
College Board
jsmith@collegeboard.org

Julia Fox
College Board
jfox@collegeboard.org

September 2014

ABSTRACT

Recent empirical work in education has demonstrated the importance both of peer effects and of various factors that affect college choices. We connect these literatures by highlighting a previously unstudied determinant of college choice, namely the college choice made by one's older sibling. Data on 1.6 million sibling pairs of SAT-takers reveals that younger and older siblings' choices are very closely related. One-fifth of younger siblings enroll in the same college as their older siblings. Conditional on their own academic skill and other characteristics, younger siblings are about 15 percentage points more likely to enroll in four-year colleges or highly competitive colleges if their older siblings do so first. These findings vary little by family income, race or proximity to four-year colleges. Younger siblings are more likely to follow the college choices of their older siblings the more they resemble each other in academic skill, age and gender. These results may improve the targeting of college choice interventions and, more importantly, should prompt further research on the sharing of information and shaping of educational preferences within families.

* This research reflects the views of the authors and not their corresponding institutions.

The truth is that if Princeton hadn't found my brother as a basketball recruit and if I hadn't seen that he could succeed on a campus like that, it never would have occurred to me to apply to that school, never.

--Michelle Obama

Introduction

For decades, researchers from various disciplines have tried to model how students make college enrollment decisions. Such disciplines include economics (Fuller, Manski and Wise, 1982), sociology (Hearn, 1991), and education (Jackson, 1978). The modeling problem has, however, proven difficult, for at least three reasons. First, there are thousands of colleges, each with numerous attributes. Second, students have heterogeneous preferences for college enrollment and for those college attributes. Third, students differ in the extent to which they have accurate information about potential colleges. Many of these factors are unobservable to the econometrician modelling college choice.

A few unsurprising characteristics of college have emerged from this literature as important to the college decision. First, the cost of college and the availability of financial aid are important factors in students' decisions, particularly for low-income students (Avery and Hoxby, 2004; Dynarski, 2003; Hurwitz, 2012). Second, proximity to colleges increases the likelihood that students enroll as students, and particularly low-income students, prefer colleges closer to home (Hossler et al., 1989; Leppel, 1993; DesJardins et al. 1999). Third, college quality has become an increasingly important determinant of students' enrollment choices (Long, 2004), with small changes in college rankings affecting the number of applicants to a given college (Luca and Smith, 2013).

Other determinants of college choice are harder to explain from a model of fully rational behavior on the part of students. High-achieving low-income students do not apply to or enroll in the same quality colleges as their higher income peers, despite the fact that the students would likely pay very little at these selective institutions (Hoxby and Avery, 2012). Many students apply only to the number of colleges for which it is free to send their test scores, such that even an elimination as small as $6 in cost can substantially change students' college choices (Pallais, 2013). Colleges receive substantially fewer applications when they increase their application fees by a few dollars or add an admission essay (Smith et al., 2014), but more applications when their sports teams succeed (Pope and Pope, 2009). Relatively small amounts of merit aid can induce students into colleges of dramatically lower quality, harming their own graduation rates (Cohodes and Goodman, 2014). The fact that

relatively small interventions, such as information mailings with application fee waivers (Hoxby and Turner, 2013), help with the completion of financial aid forms (Bettinger et al. 2012), or mandatory college entrance exams (Hurwitz et al. Forthcoming; Klasik, 2013) can increase enrollment suggests that such suboptimal behavior is likely driven by a combination of information gaps and behavioral biases (Dillon and Smith, 2013).

Given the volume of research into determinants of college choice, it is therefore remarkable that nearly nothing has been written explicitly about the influence of family members themselves. Consideration of families are, of course, implicit in much of the aforementioned research, in that most analyses control for or even estimate the impact of family factors such as parental income and education. Some papers exploit family structure in their analyses, using twin or other sibling fixed effects to account for selection bias when estimating returns to college quality (Ashenfelter and Krueger, 1994; Behrman, Rosensweig, and Taubman, 1996; Rouse, 1999; Lindahl and Regner, 2005; Smith, 2013). Yet others study birth order effects on educational attainment, though these often focus on differential sources of parental investment (Behrman and Taubman, 1986; Black, Devereaux, and Salvanes, 2005; Kantarevic and Mechoulan, 2006; Booth and Kee, 2009; Hotz and Patano, 2013).

It is perhaps even more remarkable that little has been written on the particular influence of siblings on each other's educational decisions. Though a fairly extensive literature documents sibling influences on risky behaviors such as smoking and drinking (Altonji, Cattan and Ware, 2010), we are aware of only three papers that attempt to measure the influence of siblings on each other's educational decisions. Using the NLSY79, Oettinger (2000) argues that older siblings' high school graduation status influences the high school graduation status of younger siblings, addressing endogeneity of the former by instrumenting with gender, family structure and unemployment rates. Loury (2004) estimates that, controlling for a host of other variables, African-Americans' college enrollment rates are substantially higher when they have older siblings who have enrolled in college. Using Danish data, Joensen and Nielsen (2013) show that quasi-experimental variation in older siblings' access to advanced math and science coursework alters the coursework choices of younger siblings.

Effects of other sorts of peers have, of course, been extensively documented (Sacerdote, 2011). The now vast literature on peer effects rarely considers siblings as peers, instead studying interactions between classmates, schoolmates or roommates. That literature most frequently estimates impacts of peers on student achievement or behavior, rarely if ever using college choice as an outcome. We therefore connect the literature on college choice to the literature on peer effects by carefully investigating the relationship between siblings' college enrollment decisions. To do so, we use data on

the SAT scores and college choices of the universe of SAT-takers from the 2004-11 high school graduation cohorts. Among the approximately 10 million students in those cohorts, we identify 1.6 million pairs of siblings by matching students on last names and home addresses. We then analyze simple college choice models in which the younger siblings' enrollment choices are regressed on a rich set of demographic and academic skill controls, as well as on variables measuring the college enrollment choices of their older siblings. We also explore the extent to which the relationship between siblings' college choices varies by siblings' similarities in terms of academic skill, age and gender.

We show that younger and older siblings' choices are very closely related. One-fifth of younger siblings enroll in the same college as their older siblings. Conditional on their own academic skill and other characteristics, younger siblings are about 15 percentage points more likely to enroll in four-year colleges or highly competitive colleges if their older siblings do so first. The quality of college selected by an older sibling is strongly predictive of the quality chosen by a younger sibling. These findings vary little by family income, race or proximity to four-year colleges. Younger siblings are more likely to follow the college choices of their older siblings the more they resemble each other in academic skill, age and gender. Our hope is that these results may improve the targeting of college choice interventions and, more importantly, prompt further research on the sharing of information and shaping of educational preferences within families.

We turn now to a description of the data. After that, we explain in detail how we estimate the relationship between siblings' college choices and discuss the magnitude of these estimates. We then explore whether such estimates vary by the similarity of the siblings. Finally, we discuss a number of theoretical reasons why siblings' college choices might affect each other. We argue that the relationships estimated here are at least partly causal. We conclude with implications for future work.

Data and Summary Statistics

Our data set comes from the College Board's (CB) universe of SAT-takers in the 2004-11 high school graduation cohorts. Every year, approximately 1.5 million high school students take the SAT, a standardized test often required for college admissions. The test has both a math and critical reading section, each graded on a scale of 200 to 800 for a maximum possible total score of 1600.[1] We observe

[1] In 2005, CB added a writing section. For continuity across the sample, and because admissions put most weight on the math and critical reading sections, we only consider those two sections.

those test scores, as well as a rich set of variables self-reported by students, including high school GPA, gender, race/ethnicity, parental education and parental income.[2] We also observe up to 30 Score Sends, the process by which a student requests that CB officially provide his or her SAT score to a given college. Score Sends have been shown to be good proxies for actual college applications (Card and Krueger, 2005; Pallais 2013). The CB data also include each student's full name and home address, which we use to identify siblings, as well as the high school attended.

The data on SAT takers has been merged with National Student Clearinghouse (NSC) data that tracks postsecondary enrollment information for more than 94 percent of students enrolled in U.S. postsecondary institutions.[3] Data from the NSC allow us to observe which college, if any, a student enrolls in after high school graduation. Using data from the Integrated Postsecondary Education Data System (IPEDS), we further characterize each college as a two-year or four-year institution and, for four-year colleges, by the average SAT score of incoming students. We also consider as an outcome attendance at one of the roughly 200 colleges classified as "highly competitive" based on the top two categories in the Barron's Admissions Competitive Index ("most competitive" and "highly competitive").

Students do not list siblings on their SAT registration forms. We identify each student's siblings as those who share the student's last name and home address.[4] Because the data contain eight cohorts and students generally take the SAT around age 17, sibling pairs in our data can be at most about eight years apart in age. Our method of identifying siblings is unlikely to falsely identify two students as siblings given how unusual it would be for two individuals close in age to share a last name and home address without being siblings. The method does, however, fail to match siblings if their families have changed home address between the times the two siblings took the SAT. Furthermore, we miss any siblings who graduated from high school before 2004 or after 2011. We focus on the younger sibling as the unit of analysis because we are interested in the potential influence of older siblings. Our final analytic sample consists of younger siblings from the high school classes of 2005-11 for whom we could find older siblings in the classes of 2004-10.

We present summary statistics in Table 1, where the first column contains all 10 million SAT-takers and the second column contains the 1.6 million SAT-takers we identify as younger siblings. Panels A and B list some of the demographic variables and academic skill measures we observe. About 5% of students fail to report race, nearly half fail to report parental income, and 10% fail to report parental

[2] The self-reported variables are sometimes missing. Rather than drop those observations, we generate indicators for those instances.

[3] A large fraction of non-participating colleges are for-profit institutions.

[4] For home address, we use city, state, and the first five characters of the street, including street number.

education. [5] Compared to the full population of SAT-takers, younger siblings are less likely to be black or Hispanic, to come from families earning less than $50,000 a year, and to have parents with no college education. Younger siblings have slightly higher GPAs and higher SAT scores than the overall population of SAT-takers.[6] That younger siblings are more advantaged demographically and academically is likely driven by the fact that we can only identify siblings if they come from families with multiple children taking the SAT and whose addresses are stable over time. Disadvantaged families are less likely to satisfy both of these conditions.

Panel C of Table 1 shows the college enrollment outcomes we analyze. We define all such outcomes using the college each student enrolls in the fall after their high school graduation. 73% of younger siblings enroll in four-year colleges, 21% enroll in two-year colleges, and 7% do not enroll in any college or are missing college enrollment data, perhaps by attending a college not included in the NSC. One-fifth enroll in a highly competitive college, as defined by Barron's. Finally, conditional on enrolling in a four-year college, the average SAT scores of incoming students at younger siblings' enrolled colleges is 1130.

Some of the younger siblings in our data have multiple older siblings we can identify. Among the sets of siblings identified in the data, 90% are pairs, 9.4% are triads and just under 1% are sets of four or more. To define a unique older sibling for estimation purposes, we can associate each younger sibling either with the oldest identifiable sibling or with the older sibling closest in age (i.e. the youngest older sibling). We choose to use the oldest sibling as the relevant older sibling. Though not reported here, none of our results are substantially affected by this choice, in part because so many of our younger siblings have only a single older sibling in the data.

Panel D of Table 1 shows various ways in which we can compare younger and older siblings, both in terms of demographic and academic background and in terms of college choices. Comparing siblings' characteristics allows us to explore whether the relationship between siblings' colleges choices is stronger when the siblings are more similar in terms of gender, age or academic skill. We observe, for example, that 51% of the younger siblings have the same gender as their older siblings. The average age difference between siblings in our data is 2.8 years. More than half of sibling pairs differ in their composite (M+CR) SAT scores by 100 points, or roughly half of a standard deviation. We will estimate below the extent to which sibling differences in gender, age and academic skill relate to differences in siblings' college choices.

[5] Thought not listed here, about 3% report "Other race" in both the full and younger siblings sample.
[6] The nearly 40 point difference in SAT scores represents roughly 0.2 standard deviations during this time period.

To further motivate this exploration, we note the striking results in panel D concerning the absolute proportion of younger siblings who make similar college choices as their older siblings. 31% of younger siblings apply to the college their older sibling attended, as measured by SAT Score Sends. A remarkable 19% of younger siblings enroll in the same college as their older sibling. 69% of younger siblings enroll in the same the same type of college as their older sibling, where type is defined as 4-year, 2-year or no college. Finally, 36% of younger siblings enroll in a college in the same Barron's category as their older siblings. All of this suggests that siblings' college choices are strongly related. The analysis below attempts to unpack some of the determinants of that relationship.

The Predictive Power of Older Siblings' College Choices

We begin by estimating the relationship between younger and older siblings' college choices. To do so, we run two types of OLS regressions for each college choice outcome *Y* of student *i* in high school *s* and graduation year *t*:

$$Y_{ist} = \alpha X_i + \delta_{st} + u_{ist} \quad (1)$$

$$Y_{ist} = \alpha X_i + \delta_{st} + \beta SibCollegeChoice + u_{ist} \quad (2)$$

Equation 1 is the baseline regression that includes only characteristics of the younger sibling. These include a full set of indicators for gender, race/ethnicity, parental income and parental education, as well as controls for the younger sibling's own high school GPA and SAT score.[7] We also include high school by year fixed effects, so that students are always being compared to their high school cohort peers. We cluster standard errors by family to account for the fact that some families contain more than one younger sibling in our data. Equation 2 adds a measure of the older sibling's college choice, such as whether the older sibling enrolled in a four-year college. The coefficient β describes the extent to which older siblings' college choices are related to the college choices of their younger siblings.

Table 2 shows the results of these regressions. The first two columns use as an outcome an indicator for the younger sibling's enrollment in any four-year college. Column 1 shows that, conditional on academic skill and all other variables, four-year college enrollment rates are 4.4 percentage points lower for men than for women, 10.5 percentage points higher for black students than for white students, and 3-7 percentage points lower for students from lower income and less educated families.

[7] When these variables are missing, a dummy variable is created to indicate as such.

Unsurprisingly, both measures of a student's own academic skill, high school GPA and SAT score, strongly predict four-year college enrollment. Remarkably, these fairly rich controls for student background and academic skill, as well as high school-by-year fixed effects, explain only 29.3 percent of the variation in four-year college enrollment.

In column 2, adding an indicator for whether the older sibling enrolled in a four-year college generates three important results. First, older siblings' four-year college enrollment strongly predicts younger siblings' four-year college enrollment. Conditional on their own characteristics and academic skill, and compared to their high school cohort peers, younger siblings are a substantial 14.6 percentage points more likely to enroll in a four-year college if their older sibling did so. Second, adding that single variable increases the fraction of variation in four-year college enrollment explained by 1.6 percentage points, to 30.9 percent. Third, adding the older sibling's college choice does not substantially change the coefficients in panel A, indicating that students' own characteristics and older siblings' college choices are independent predictors of college choice. Finally, though not shown here, including an additional control for older siblings' SAT score has little impact on these conclusions.

We obtain similar results if we instead use measures of college quality as outcomes. The estimates in column 4 suggest that, conditional on all else, a younger sibling is 17.1 percentage points more likely to enroll in a highly competitive college if his or her older sibling did so. In columns 5 and 6, we estimate the relationship between the quality of siblings' chosen colleges as measured by institution-level SAT scores.[8] The relevant coefficient implies that an older sibling enrolling in a college with an average SAT score 100 points higher is associated with an increase of 13.5 SAT points in the average score of the younger sibling's chosen college. This strong relationship between the quality of siblings' college choices can be seen in the unconditional data shown in Figure 1, which simply plots the relationship between the institution-level SAT scores of siblings' chosen colleges.[9]

In Table 3, we perform similar analyses as in Table 2, but use as outcomes measures of the match quality between younger siblings and their chosen colleges. We first define a measure of match quality as the difference between a student's own SAT score and the average SAT score of incoming students at the college he or she enrolls in. Students for whom this measure is large and positive have

[8] In some instances, one or both siblings did not have values for the average SAT score of first four-year college attended, either because the sibling lacked enrollment data, attended a two-year college or attended a four-year college without IPEDS-reported average SAT scores. We assigned those missing values an average SAT of college enrolled the lowest average SAT of 675 and separately included dummies for whether this value was assigned to the younger sibling and/or older sibling.

[9] For graphical purposes, Figure 1 excludes the observations where either a younger or older sibling has no reported average SAT of college enrolled.

college peers much less academically skilled than they are and are thus, by definition, "undermatched". Students for whom this measure is large and negative have college peers much more academically skilled than they are and are thus "overmatched".[10] To operationalize this, we define indicators for the difference between a student's SAT score and the college's SAT score being greater than 100 points (undermatch) or less than -100 points (overmatch).[11]

Columns 1 and 2 use the continuous measure of match quality, which we refer to as the extent of undermatch because positive coefficients imply increased undermatch. The strong relationship between siblings' extent of undermatch is shown in Figure 2, which plots the unconditional relationship between these variables. The extent of the older sibling's undermatch is clearly and positively related to the extent of the younger sibling's undermatch. Conditional on all else, each 100 SAT point increase in the extent of the older sibling's undermatch is associated with a 3.3 SAT point increase in the extent of the younger sibling's undermatch.

We see similar results using the binary measures of undermatch and overmatch as outcomes. The estimates in columns 4 and 6 suggest that, conditional on all else, a younger sibling is a substantial four percentage points more likely to undermatch if his or her older sibling undermatched. The relationship between siblings' overmatch probabilities is nearly identical. These coefficients are larger in magnitude than any of the other binary demographic variables, suggesting that the sibling "gap" in match quality exceeds any observed gender, race or income gaps. In short, all three measures suggest that older siblings' college match quality is strongly predictive of the college match quality of their younger siblings.

In Table 4, we explore whether the observed relationships between younger and older siblings' college choices vary by the characteristics of the families from which they come. To do so, we re-run the regressions from the even-numbered columns of Tables 2 and 3, separating the sample by income, race, and the density of nearby colleges. For each regression, we then list in Table 4 only the coefficient measuring the relationship between siblings' college choices (i.e. the one listed in panel B in Tables 2 and 3). Columns 1 and 2, which divide the sample into families reporting less than and more than $50,000 in annual income, show no substantial differences in the magnitudes of the estimated sibling relationships. Columns 3 and 4, which divide the sample into black and Hispanic families and white families, similarly show no systematic or large differences by race in these relationships. In columns 5 and 6, we compute the mean number of four-year colleges within a 25 mile radius of home (10.3

[10] This undermatch and overmatch terminology first appears in Bowen, Chingos, and McPherson (2009).

[11] Again, we assign students with missing average SAT of college enrolled a 675 and include corresponding dummy variables to indicate as such.

colleges) for students in our sample. We then divide the sample into students living below and above that mean, so that the former are living near relatively few four-year colleges and that latter are living near relatively many four-year colleges. As with income and race, there is little evidence of substantial heterogeneity by proximity to four-year colleges.

Do Similar Siblings Have Stronger College Choice Relationships?

Having documented the strong relationship between the college choices of younger and older siblings, we now explore whether siblings' tendency to emulate each other depends upon the extent to which the siblings are similar in terms of academic skill, gender and age. We do this by running regressions of the form:

$$Y_{ist} = \alpha X_i + \delta_{st} + \beta_1 SibSAT + \beta_2 SkillDiff + \beta_3 GenderDiff + \beta_6 AgeDiff + u_{ist} \quad (3)$$

Here, the outcome *Y* indicates whether the younger and older sibling have made the same college choice, such as applying to the same college, enrolling in the same college, enrolling in the same type of college (4-year/2-year/none) or enrolling in a college in the same Barron's category. As before, we control for characteristics of the younger sibling, including a full set of indicators for race/ethnicity, parental income and parental education, as well as the younger sibling's own high school GPA and SAT score. In addition to the younger sibling's skill measures, we control for the older sibling's SAT score as well as *SkillDiff*, an indicator for whether the absolute value of the difference between the siblings' SAT scores is greater than 100.[12] The latter measures whether the siblings differ substantially in terms of academic skill. We include *GenderDiff,* an indicator for the two sibling's having different genders, and *AgeDiff*, which measures the siblings' age difference in years.[13] As before, high school-by-year fixed effects are included to compare younger siblings to their own high school classmates.

Table 5 shows the results of these regressions, with panel A showing gender and family background factors shared by siblings (race, parental income and education), as well as controls for own and older sibling's academic skills. The first three rows imply that, conditional on all else and relative to

[12] Other functional forms of the academic difference were tested. This parameterization highlights the importance of large differences in skills between siblings.

[13] We observe few consistent differences between sister-sister and brother-brother pairs, so that this specification is the simplest that captures the interesting features of the data.

white students, black students are one percentage point less likely to apply to and six percentage points less likely to enroll in the same colleges as their older siblings. Hispanic students are two percentage points more likely to follow their older siblings to college. Asian students are eight percentage points more likely to apply to and three percentage points more likely to enroll in the same colleges as their older siblings. Relative to students from high income and B.A.-holding families, those from lower income and less educated families are slightly less likely to apply to but slightly more likely to enroll in the same colleges as their older siblings. Disadvantaged students are also less likely to mimic their older siblings' choice of college type (four-year, two-year or none) but there are no differences by income or parental education in the probability of enrolling in the same Barron's category of college as one's older sibling. The last three rows show that higher academic skills, particularly high school GPA, predict a higher likelihood of following an older sibling's college choices.

Panel B highlights potential differences between siblings, such as academic skill, gender and age. The first row implies that, conditional on the academic skills of each sibling controlled for in panel A, younger siblings whose SAT scores differ by more than 100 points from the scores of their older siblings are six percentage points less likely to apply to and enroll in the same colleges as those older siblings. They are 10 percentage points less likely to enroll in a college of the same Barron's category as their older sibling. The second rows implies that, across all of these outcomes, having an older sibling of the opposite gender decreases by one to three percentage points the probability of following that sibling's college choice. The last row of panel B implies that each year of age difference between siblings reduces by 0.6-0.8 percentage points the likelihood of a younger sibling following an older sibling's college choice. This effect is relatively linear across the distribution of age differences observed in our data. Taken as a whole, these results suggest that the more similar siblings are to each other, in terms of academic skill, gender and age, the more likely a younger sibling is to make the same college decisions as his or her older sibling. This is consistent with Joensen and Nielsen (2013), who also find that siblings close in age and of the same gender have stronger influences over each other.

Discussion and Conclusion

This paper documents a number of previously unknown facts about intra-family patterns in college enrollment. First, many younger siblings apply to and enroll in the same colleges as their older siblings. Second, even controlling for a rich set of covariates, older siblings' college enrollment decisions are strongly predictive of their younger siblings' decisions about whether to enroll and which quality of college to enroll in. Third, these strong relationships between siblings' college choices vary little by

income, race or proximity to four-year colleges. Fourth, younger siblings are more likely to follow the college choices of their older siblings the more they resemble each other in terms of academic skill, age and gender.

These facts, taken as a whole, are consistent with the possibility that the college decisions of older siblings influence the college decisions of younger siblings. There is, however, a potential non-causal explanation for these patterns, namely that the covariates available to us for this analysis are insufficient to control for fundamental differences between families that determine college enrollment choices. These could include differences in educational preferences, information about college or the labor market, or access to credit. It may be that siblings simply have the same preferences for factors such as college quality and distance from home that result from a shared environment. If the available covariates do not completely absorb such inter-family differences, then the strong relationship between siblings' choices may be partly picking up those unobserved differences. If so, older siblings' college choices reveal something about a family's type, in which case it is unsurprising that such choices then predict those of younger siblings.

There are, however, at least five channels through which older siblings' college choices might causally influence their younger siblings. First, an older sibling's application and enrollment experiences may provide information to the younger sibling that would otherwise have been costly or even impossible to obtain. This could include information about the application process, the probability of admission, the net price of enrollment, or the quality of the actual enrollment experience. Recent evidence that informational interventions (Hoxby and Turner, 2013) and exposure to teachers or schoolmates who attended selective colleges (Bulman et al., 2014; Hoxby and Avery, 2012) may affect college application and enrollment choices strongly suggests that students often operate without full information about available college options. Older siblings may thus lower the cost of obtaining such information in a classical choice model or, in a more behavioral model of choice, may increase the salience of a given college or set of colleges. The relevant information here may be about the college the older sibling enrolls in or may be about colleges more generally. Second, a younger sibling may derive specific benefits from being enrolled at the same college as an older sibling. Siblings may enjoy each other's company and thus derive utility from being on the same campus. They may study together or help each other with course selection, improving their educational experience. They may live or carpool together, reducing the overall costs of enrollment. The third channel involves college pricing. Federal financial aid is, for example, calculated in part based on the number of children in the family currently enrolled in college. Having an older sibling enrolled in any college can therefore increase the amount aid

for which a younger sibling is eligible, reducing the net price of college. Younger siblings may also benefit from the fact that some colleges offer tuition discounts if multiple siblings attend the same campus.[14] A fourth channel worth mentioning comes from the fact that some colleges are more likely to admit legacy students than observationally similar non-legacy students (Hurwitz, 2011). At such colleges, an older sibling's enrollment often qualifies a younger sibling as such a legacy student.

The first two channels mentioned above, namely lowered information costs and the impacts of having an older sibling on the same campus, could in theory either raise or lower the probability of a younger sibling following an older sibling's decision. New information could reveal positive or negative facets of the application or enrollment process. Younger siblings might benefit from having older siblings on campus or might find that proximity unpleasant. Conversely, the price and legacy channels have a theoretically unambiguous sign, as they can only serve to increase the probability of a younger sibling following an older sibling's choice. Finally, a fifth channel, namely the wealth effect of having an older sibling enrolled in college, unambiguously implies a negative impact on the probability of the younger sibling's enrollment. Given that some families are likely credit constrained with respect to the total budget they can spend on their children's college education, an older sibling's enrollment may exhaust funds that would otherwise have been able to finance a younger sibling's enrollment.

In short, there are a number of theoretical reasons why we might expect to find a causal impact of an older sibling's college choice on a younger sibling. Though we cannot rule out the possibility that the relationships we have documented are simply correlational, two sets of facts documented here suggest that the non-causal story cannot fully explain the strength of the observed correlations. First, the lack of heterogeneity in these sibling relationships by proximity to four-year colleges is inconsistent with the possibility that inter-family differences in geographic access to colleges is driving our results. This rules out the explanation that students simply choose the nearest college and that, for siblings, that nearest college is always the same. The lack of heterogeneity by income and race also make it less likely that these relationships are purely about unobserved inter-family differences, as the magnitude of those unobserved differences would need to be quite similar across income and race categories to generate results of such similar magnitudes.

[14] Though we cannot find any definitive estimate of the number of colleges that offer sibling discounts, multiple recent news articles suggest the practice is fairly common. For example, see U.S. News and World Report's "Some Colleges Discount Tuition for Siblings", by Katy Hopkins, December 11, 2012, available at http://www.usnews.com/education/best-colleges/paying-for-college/articles/2012/12/11/some-colleges-discount-tuition-for-siblings.

Second, the fact that younger siblings more strongly follow the choices of more similar older siblings seems difficult to attribute to unobserved inter-family differences. In particular, the fact that being of the same gender increases the probability of following an older sibling's college choice is inconsistent with that story, given the low probability that the gender composition of siblings is highly correlated with inter-family differences in educational preferences (i.e. conditional on having two children, the gender of each is random). That siblings closer in age are more likely to make similar college choices is also difficult to explain away with such unobserved family factors. A much more likely explanation is that younger siblings of the same gender as or close in age to their older siblings are more likely to value the older siblings' information about the college application process, information about the experience of attending a particular college, or physical proximity on campus. All of these possibilities suggest a causal influence of the older sibling's college choices.

Ultimately, the estimates presented in this paper cannot rule out the possibility that older siblings' college choices are simply proxies for unobserved differences between families. If so, then at the very least this suggests that such differences are not well-captured by existing variables available to researchers in most administrative data sets, itself an interesting fact. It also suggests that older siblings' college choices can be used to help better predict the choices of younger siblings. Numerous recent interventions, such as Hoxby and Turner (2013), have been targeted at "at-risk" students, but identifying exactly which students are most at risk poses a challenge if model explanatory power is weak. We find that 37% of those we predict to undermatch using our basic regression model (1) are type I errors, or those who turn out not to undermatch. Including an indicator for older sibling's undermatch, as in regression model (2) reduces that error rate to 33%. We therefore have some evidence that high school counselors and schools might find it worthwhile to flag for extra attention students whose older siblings have made apparently poor college decisions.

More work is, however, needed to establish the extent to which the college choices of older siblings influence the choices of their younger siblings. The primary empirical challenge is finding an exogenous source of variation in the college choices of older siblings. Researchers have begun to find sources of exogenous variation driving students' enrollment choices generally (Goodman et al., 2014). Modern administrative data sets, which are large and rich with detail, make it increasingly likely that such variation can be combined with information on family structure to identify the causal impact of older siblings' choices. Identifying which channels, if any, are responsible for such impacts will be crucial to making such research findings useful to policymakers.

References

Altonji, J. G., Cattan, S., and I. Ware. (2010). "Identifying sibling influence on teenage substance use," National Bureau of Economic Research, Working Paper #16508.

Ashenfelter, O. and A. Krueger (1994). "Estimates on the Return to Schooling From a New Sample of Twins," *American Economic Review*. 84: 1157-1173.

Avery, C. and C. Hoxby (2004). "Do and Should Financial Aid Decisions Affect Students' College Choices?" in Caroline Hoxby, ed. *College Choices: The New Economics of Choosing, Attending, and Completing College*. University of Chicago Press.

Behrman, J., M. Rosensweig, and P. Taubman (1996). "College Choice and Wages: Estimates Using Data on Female Twins," *The Review of Economics and Statistics*. 78(4): 672-685.

Behrman, J., and P. Taubman (1986). "Birth Order, Schooling, and Earnings," *Journal of Labor Economics*. 4: S121-45.

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

Black, S, P. Devereaux, and K.Salvanes (2005). "The More the Merrier? The Effect of Family Size and Birth Order on Children's Education." *Quarterly Journal of Economics*. 120(2): 669-700.

Booth, A. and H.J. Kee (2009). "Birth order matters: the effect of family size on educational attainment". *Journal of Population Economics*. 22, 367-397.

Bowen, W.G., Chingos, M., and McPherson, M. (2009). *Crossing the finish line*. Princeton, NJ: Princeton University Press.

Bulman, G., C. Hoxby, and J. Meer (2014). “Teach For America and Students’ College Application Choices,” Working paper.

Card, D. and A. Krueger (2005). ”Would the Elimination of Affirmative Action Affect Highly Qualified Minority Applicants? Evidence From California and Texas,” *Industrial and Labor Relations Review*, 56: 416-434.

Cohodes, S. and J. Goodman (2014). “Merit Aid, College Quality and College Completion: Massachusetts' Adams Scholarship as an In-Kind Subsidy,” *American Economic Journal: Applied Economics*.

DesJardins, S., H. Dundar, and D. Hendel (1999). Modeling the college application decision process in a land-grant university, “ *Economics of Education Review*, “ 18: 117-132.

Dillon, E. and J. Smith (2013). “The determinants of mismatch between students and colleges,” NBER Working Paper #19286.

Dynarski, S. (2003). "Does Aid Matter? Measuring the Effect of Student Aid on College Attendance and Completion," *American Economic Review*, 93(1): 279-288

Fuller, W. C., Manski, C. F., and Wise, D. A. (1982). New evidence on the economic determinants of postsecondary schooling choices. *Journal of Human Resources*, 17(4): 477–498.

Goodman, J., Hurwitz, M. and Smith, J. (2014). College Access, Initial College Choice and Degree Completion. Harvard Kennedy School Faculty Research Working Paper Series, RWP14-030.

Hearn, J. (1991). “Academic and nonacademic influences on the college destinations of 1980 high school graduates,” *Sociology of Education*. 64(3): 158-171.

Hossler, D., Braxton, J. and Coppersmith, G. (1989) “Under-standing student college choice,” In *Higher Education: Handbook of Theory and Research* (Edited by J. C. Smart), Vol. 5, pp. 231–288. New York: Agathon Press.

Hotz, J. and J. Patano (2013). “Strategic Parenting, Birth Order and School Performance,” NBER Working Paper #19542.

Hoxby, C., and C. Avery (2012). “The Missing ‘One-Offs’: The Hidden Supply of High-Achieving, Low Income Students,” NBER Working Paper No. w18586.

Hoxby, C., and S. Turner (2013) “Expanding College Opportunities for High-Achieving, Low Income Students,” SIEPR Discussion Paper No. 12-014.

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. (2012). “The impact of institutional grant aid on college choice,” *Educational Evaluation and Policy Analysis*. 34(3): 344-363.

Hurwitz, M., J. Smith, S. Niu, and J. Howell (*Forthcoming*). “ The Maine Question: How is Four-Year College Enrollment Affected by Mandatory College Entrance Exams?” *Educational Evaluation and Policy Analysis*.

Jackson, G. (1978). “Financial aid and student enrollment,” *Journal of Higher Education*. 49(6): 548-574.

Joensen, J. and H. Nielsen (2013). “Peer Effects in Math and Science.” Working Paper.

Kantarevic, J., Mechoulan, S. (2006). “Birth order, educational attainment, and earnings: an investigation using the PSID,” *Journal of Human Resources*. 41, 755–777.

Klasik, D. (2013). “The ACT of enrollment: The college enrollment effects of required college entrance exam taking.” *Educational Researcher,* 42, 151–160.

Leppel, K. (1993). “Logit estimation of a gravity model of the college enrollment decision,” *Research in Higher Education*, 34(3): 387–398.

Lindahl, L., H. Regner. (2005). “College Choice and Subsequent Earnings: Results Using Swedish Sibling Data,” *Scandinavian Journal of Economics*, 107(3): 437-457.

Long, B. (2004). “How have college applications changed over time? An application of the conditional logistic choice model,” *Journal of Econometrics*. 121: 271-296.

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

Luca, M. and J. Smith (2013). “Salience in Quality Disclosure: Evidence from the U.S. News College Rankings,” *Journal of Economics and Management Strategy*, 22(1): 58-77.

Oettinger, G. (2000). “Sibling similarity in high school graduation outcomes: Causal interdependency or unobserved heterogeneity?,” *Southern Economic Journal*: 631-648.

Pallais, A. (2013). “Small Differences that Matter: Mistakes in Applying to College,“ NBER Working Paper #19480.

Pope, D. and J. Pope (2009). “The Impact of College Sports Success on the Quantity and Quality of Student Applications,” *Southern Economic Journal*. 75(3): 750-780.

Rouse, C. (1999). “Further Estimates of the Economic Return to Schooling From a New Sample of Twins,” *Economics of Education Review*. 18(2): 149-157.

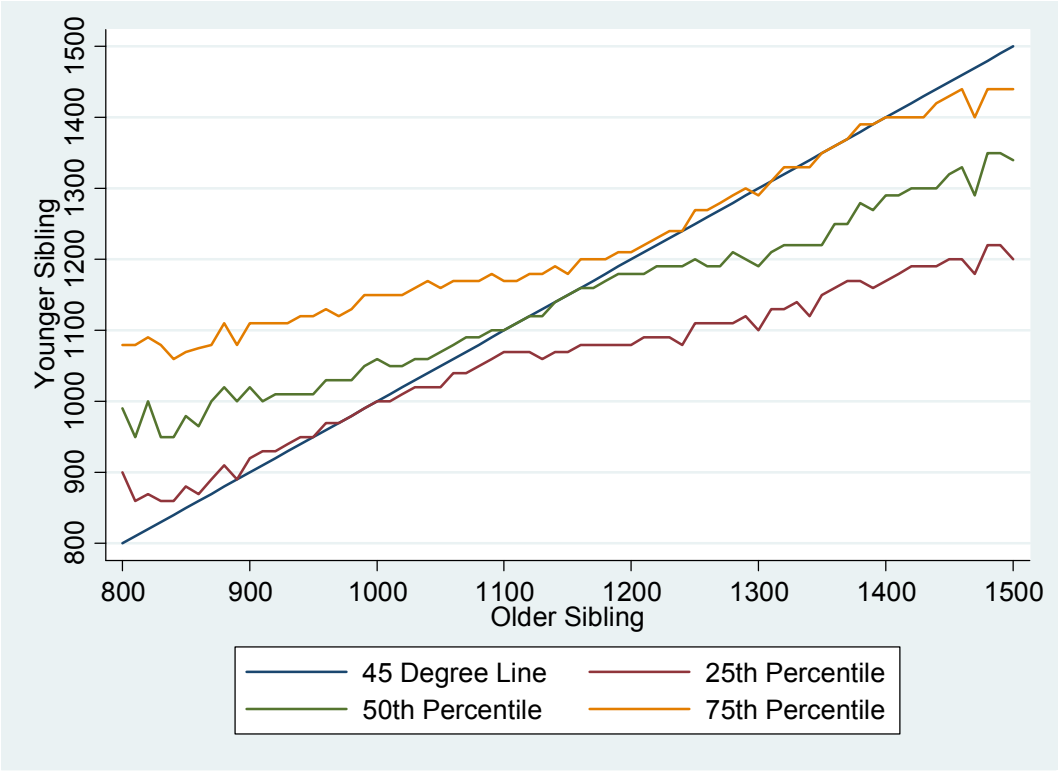
Sacerdote, B. (2011). “Peer effects in education: How might they work, how big are they and how much do we know thus far?” *Handbook of the Economics of Education*, 3, 249-277.

Smith, J. (2013). “Ova and Out: Using Twins to Estimate the Educational Returns to Attending a Selective College,” *Economics of Education Review*, 36: 166-180.

Smith, J., M. Hurwitz, and J. Howell (2014). “Screening Mechanisms and Student Responses in the College Market,” Working paper.

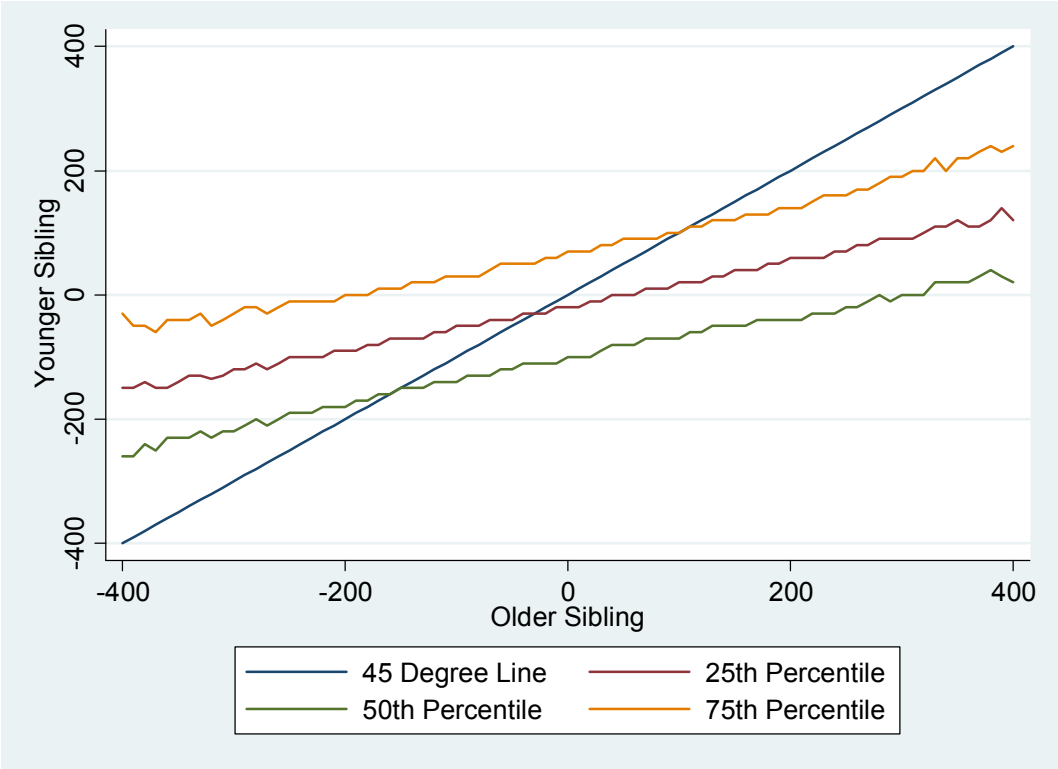
Figures and Tables

Figure 1 - Average SAT of Siblings' Enrolled Colleges



Notes: Includes all identified siblings who took the SAT between 2004 and 2011. Only includes younger and older sibling pairs who both attend a four-year college with reported average SAT scores. The older siblings college's average SAT is rounded to the nearest ten SAT points and are plotted on the x-axis. For all older siblings with a given average college SAT score, the 25th, 50th, and 75th percentiles of younger sibling's average SAT of enrolled colleges are calculated.

Figure 2 - Difference in Siblings' SATs and Average SATs of Enrolled Colleges



Notes: Includes all identified siblings who took the SAT between 2004 and 2011. Only includes younger and older sibling pairs who both attend a four-year college with reported average SAT scores. The older siblings college's average SAT is rounded to the nearest ten SAT points, subtracted from their own SAT score, and are plotted on the x-axis. For all older siblings with a given average difference, the 25th, 50th, and 75th percentiles of younger sibling's differences are calculated.

Table 1 - Summary Statistics

	All students (1)	Younger siblings (2)
(A) Demographic characteristics		
Female	0.538	0.522
White	0.573	0.682
Black	0.119	0.065
Hispanic	0.128	0.093
Asian	0.087	0.087
Parental income: $0-$50,000	0.193	0.109
Parental income: $50,000-$100,000	0.208	0.198
Parental income: $100,000+	0.152	0.204
Parental income: Missing	0.447	0.490
Parental education: High school or less	0.164	0.108
Parental education: Some college or A.A.	0.245	0.201
Parental education: B.A. or higher	0.490	0.603
Parental education: Missing	0.102	0.087
(B) Academic skill		
High school GPA[1]	3.323	3.390
SAT score (Math + Critical Reading)	1016	1055
(C) College enrollment		
Four-year college	0.648	0.725
Two-year college	0.255	0.206
No college	0.097	0.069
Highly competitive college (Barrons' s Most or Highly Compeiti	0.159	0.195
SAT score of enrolled college[2]	1118	1130
(D) Relation to older sibling		
Same gender		0.510
Age difference		2.791
|Own SAT - older sibling SAT| > 100 SAT Points		0.563
Applies to older sibling's college[3]		0.309
Enrolls in same college[4]		0.187
Enrolls in same type (4-year/2-year/none)		0.687
Enrolls in same Barron's category[4]		0.358
N	10,044,488	1,614,007

Notes: The sample in column 1 includes all SAT test-takers from the high school graduation cohorts of 2005-2011. Column 2 includes only students who could be identified as younger siblings. Demographic characteristics and high school GPA are self-reported by students to the College Board. College enrollment outcomes come from the National Student Clearinghouse.

1. Among students with reported GPA (94%).
2. Among students who enroll in college with reported SAT.
3. Equals zero if older sibling did not enroll in college.
4. Equals zero if younger or older sibling did not enroll in college.

Table 2 - The Relationship Between Older and Younger Siblings' College Choices

	Enrolls in four-year college		Enrolls in highly competitive college		Average SAT score of enrolled college	
	(1)	(2)	(3)	(4)	(5)	(6)
(A) Younger sibling's characteristics						
Male	-0.044***	-0.045***	-0.016***	-0.015***	-1.054***	-0.848***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.137)	(0.135)
Black	0.105***	0.103***	0.072***	0.072***	1.987***	3.995***
	(0.002)	(0.002)	(0.001)	(0.001)	(0.372)	(0.367)
Hispanic	-0.006***	-0.000	0.036***	0.034***	8.184***	7.714***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.281)	(0.277)
Asian	-0.001	-0.002	0.078***	0.064***	15.020***	12.659***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.290)	(0.284)
Parental income: $0-$50,000	-0.036***	-0.030***	-0.013***	-0.008***	-2.800***	-1.606***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.264)	(0.261)
Parental income: $50,000-$100,000	-0.032***	-0.027***	-0.027***	-0.020***	-5.959***	-4.574***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.203)	(0.201)
Parental education: High school or less	-0.069***	-0.054***	-0.008***	0.002**	-2.016***	-0.033
	(0.001)	(0.001)	(0.001)	(0.001)	(0.231)	(0.229)
Parental education: Some college or A.A.	-0.055***	-0.043***	-0.025***	-0.014***	-6.573***	-4.263***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.170)	(0.168)
High school GPA	0.153***	0.151***	0.105***	0.103***	23.047***	22.706***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.140)	(0.138)
SAT score	0.057***	0.053***	0.072***	0.064***	17.564***	15.964***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.052)	(0.051)
(B) Older sibling's SAT score and college choice						
Older sibling enrolls in four-year college	--	0.146***	--	--	--	--
	--	(0.001)	--	--	--	--
Older sibling enrolls in highly competitive college	--	--	--	0.171***	--	--
	--	--	--	(0.001)	--	--
Average SAT of older sibling's enrolled college	--	--	--	--	--	0.135***
	--	--	--	--	--	(0.001)
R-squared	0.293	0.309	0.351	0.375	0.902	0.904

Notes: Standard errors are clustered at the family level and in parentheses (* p<.10 ** p<.05 *** p<.01). The sample includes all SAT test-takers from the high school graduation cohorts of 2005-2011 who could be identified as younger siblings (N = 1,614,007). Older siblings are defined as the oldest sibling observed in the data. In columns 1-2, the outcome is an indicator for the younger sibling enrolling in any four-year college. In columns 3-4, the outcome is an indicator for enrollment in a college ranked as highly competitive by Barron's. In columns 5-6, the outcome is the average SAT score of the four-year college enrolled in (missing avg. SAT and non-enrollee are assigned the lowest avg SAT of 675 and a dummy variable is included as independent variable to indicate as such). All regressions include high school by year fixed effects. Also included but not shown are indicators for other race, missing race, missing parental income, missing parental education and missing gender. Omitted categories are white, parental income greater than $100,000, parental education B.A. or higher, and female.

Table 3 - The Relationship Between Older and Younger Siblings' College Match Quality

	Younger sibling's SAT score - enrolled college SAT score		Undermatches (own SAT - college SAT > 100)		Overmatches (college SAT - own SAT > 100)	
	(1)	(2)	(3)	(4)	(5)	(6)
(A) Younger sibling's characteristics						
Male	1.054***	1.466***	0.004***	0.004***	-0.007***	-0.008***
	(0.137)	(0.137)	(0.001)	(0.001)	(0.001)	(0.001)
Black	-1.987***	0.293	-0.019***	-0.014***	0.012***	0.008***
	(0.372)	(0.372)	(0.001)	(0.001)	(0.001)	(0.001)
Hispanic	-8.184***	-7.036***	-0.014***	-0.012***	0.011***	0.009***
	(0.281)	(0.281)	(0.001)	(0.001)	(0.001)	(0.001)
Asian	-15.020***	-14.446***	-0.029***	-0.028***	0.025***	0.023***
	(0.290)	(0.289)	(0.001)	(0.001)	(0.001)	(0.001)
Parental income: $0-$50,000	2.800***	2.777***	0.010***	0.009***	-0.008***	-0.008***
	(0.264)	(0.264)	(0.001)	(0.001)	(0.001)	(0.001)
Parental income: $50,000-$100,000	5.959***	5.708***	0.017***	0.016***	-0.009***	-0.008***
	(0.203)	(0.203)	(0.001)	(0.001)	(0.001)	(0.001)
Parental education: High school or less	2.016***	2.931***	0.011***	0.012***	-0.016***	-0.016***
	(0.231)	(0.232)	(0.001)	(0.001)	(0.001)	(0.001)
Parental education: Some college or A.A.	6.573***	7.006***	0.022***	0.022***	-0.015***	-0.015***
	(0.170)	(0.170)	(0.001)	(0.001)	(0.001)	(0.001)
High school GPA	-23.047***	-22.463***	-0.056***	-0.055***	0.058***	0.057***
	(0.140)	(0.140)	(0.001)	(0.001)	(0.001)	(0.001)
SAT score	82.436***	81.442***	0.119***	0.117***	-0.091***	-0.089***
	(0.052)	(0.054)	(0.000)	(0.000)	(0.000)	(0.000)
(B) Older sibling's SAT score and college match quality						
Older sibling's SAT score - enrolled college SAT score	--	0.033***	--	--	--	--
	--	(0.000)	--	--	--	--
Older sibling undermatches	--	--	--	0.040***	--	--
	--	--	--	(0.001)	--	--
Older sibling overmatches	--	--	--	--	--	0.045***
	--	--	--	--	--	(0.001)
R-squared	0.896	0.896	0.649	0.650	0.354	0.356

Notes: Standard errors are clustered at the family level and in parentheses (* p<.10 ** p<.05 *** p<.01). The sample includes all SAT test-takers from the high school graduation cohorts of 2005-2011 who could be identified as younger siblings (N = 1,614,007). Older siblings are defined as the oldest sibling observed in the data. The average SAT of college enrolled in is assigned a 675 when missing (colleges that don't report avg. SAT and non-enrollee) and a dummy variable is included as independent variable to indicate as such. In columns 1-2, the outcome is the difference between the younger's siblings SAT score and the average SAT score of the college he or she enrolls in. In columns 3-4, the outcome is an indicator for undermatch, defined as a student's SAT score exceeding the average SAT score of his or her enrolled college by over 100 points. In columns 5-6, the outcome is an indicator for overmatch, defined as the average SAT score of a student's enrolled college exceeding his or her own SAT score by over 100 points. All regressions include high school by year fixed effects. Also included but not shown are indicators for other race, missing race, missing parental income, missing parental education and missing gender. Omitted categories are white, parental income greater than $100,000, parental education B.A. or higher, and female.

Table 4 - Heterogeneous Effects

Y = Younger sibling's college choice	Income Less Than $50k (1)	Income More Than $50k (2)	Black or Hispanic (3)	White (4)	Fewer four-year colleges within 25 miles (5)	More four-year colleges within 25 miles (6)
Older sibling's college choice						
Enrolls in four-year college	0.145***	0.140***	0.138***	0.142***	0.148***	0.139***
	(0.003)	(0.001)	(0.002)	(0.001)	(0.001)	(0.002)
Enrolls in highly competitive college	0.154***	0.166***	0.145***	0.171***	0.179***	0.153***
	(0.004)	(0.001)	(0.003)	(0.001)	(0.001)	(0.002)
Average SAT score of enrolled college	0.116***	0.123***	0.122***	0.130***	0.138***	0.125***
	(0.003)	(0.001)	(0.002)	(0.001)	(0.001)	(0.001)
Younger sibling's SAT - enrolled college SAT	0.026***	0.035***	0.028***	0.033***	0.033***	0.030***
	(0.002)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Undermatches (own SAT - college SAT > 100)	0.025***	0.036***	0.033***	0.039***	0.039***	0.041***
	(0.003)	(0.001)	(0.002)	(0.001)	(0.001)	(0.001)
Overmatches (college SAT - own SAT > 100)	0.047***	0.040***	0.047***	0.041***	0.045***	0.045***
	(0.003)	(0.001)	(0.002)	(0.001)	(0.001)	(0.001)

Notes: Each coefficient is from a separate regression. Standard errors are clustered at the family level and in parentheses (* p<.10 ** p<.05 *** p<.01). The sample includes all SAT test-takers from the high school graduation cohorts of 2005-2011 who could be identified as younger siblings and self-reported income, race/ethnicity, or zip code. Older siblings are defined as the oldest sibling observed in the data. All regressions include high school by year fixed effects. Also included but not shown are indicators for parental income and education, gender, race, high school GPA, and SAT score. Columns 5 and 6 divide the sample into families living below and above the mean number of four-year colleges within a 25 mile radius.

Table 5 - Sibling Skill, Gender and Age Differences

	Applies to same college as older sibling (1)	Enrolls in same college as older sibling (2)	Enrolls in same type (4-year/2-year/none) (3)	Enrolls in same Barron's category (4)
(A) Family background, academic skill				
Male	0.014***	0.010***	-0.015***	-0.002**
	(0.001)	(0.001)	(0.001)	(0.001)
Black	-0.014***	-0.059***	0.057***	-0.031***
	(0.002)	(0.001)	(0.002)	(0.002)
Hispanic	0.017***	0.015***	0.012***	0.008***
	(0.002)	(0.001)	(0.002)	(0.002)
Asian	0.081***	0.034***	0.006***	0.016***
	(0.002)	(0.001)	(0.002)	(0.002)
Parental income: $0-$50,000	-0.004**	0.016***	-0.032***	0.001
	(0.002)	(0.001)	(0.002)	(0.002)
Parental income: $50,000-$100,000	-0.007***	0.009***	-0.029***	-0.001
	(0.001)	(0.001)	(0.001)	(0.001)
Parental education: High school or less	-0.017***	0.019***	-0.057***	-0.002
	(0.001)	(0.001)	(0.001)	(0.001)
Parental education: Some college or A.A.	-0.023***	0.010***	-0.051***	-0.001
	(0.001)	(0.001)	(0.001)	(0.001)
High school GPA	0.052***	0.024***	0.096***	0.038***
	(0.001)	(0.001)	(0.001)	(0.001)
Own SAT score	0.010***	-0.006***	0.018***	-0.000
	(0.000)	(0.000)	(0.000)	(0.000)
Older sibling's SAT score	0.031***	-0.005***	0.030***	-0.002***
	(0.000)	(0.000)	(0.000)	(0.000)
(B) Differences between siblings				
|Own SAT - older sibling SAT| > 100	-0.063***	-0.057***	-0.054***	-0.095***
	(0.001)	(0.001)	(0.001)	(0.001)
Siblings' gender differs	-0.025***	-0.024***	-0.011***	-0.023***
	(0.001)	(0.001)	(0.001)	(0.001)
Age difference	-0.007***	-0.008***	-0.005***	-0.008***
	(0.000)	(0.000)	(0.000)	(0.000)
R-squared	0.160	0.123	0.169	0.095

Notes: Standard errors are clustered at the family level and in parentheses (* p<.10 ** p<.05 *** p<.01). The sample includes all SAT test-takers from the high school graduation cohorts of 2005-2011 who could be identified as younger siblings (N = 1,614,007). Older siblings are defined as the oldest sibling observed in the data. Each column represents a single regression, where the outcome is an indicator for the younger sibling making the same college choice as the older sibling. All regressions include high school by year fixed effects. Also included but not shown are indicators for other race, missing race, missing parental income, missing parental education and missing gender. Omitted categories are white, parental income greater than $100,000, parental education B.A. or higher, and younger sisters with older brothers.