NBER WORKING PAPER SERIES

INTERGENERATIONAL AND SIBLING PEER EFFECTS IN HIGH SCHOOL MAJORS

Gordon B. Dahl
Dan-Olof Rooth
Anders Stenberg

Working Paper 27618
http://www.nber.org/papers/w27618

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
July 2020, June 2021

We thank colleagues and seminar participants at several universities and conferences for valuable feedback and suggestions. We are grateful to Hans Eric Ohlson at Statistics Sweden for help in locating and accessing the data. The project received generous financial support from the Institute for Social Research at Stockholm University. The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

Intergenerational and Sibling Peer Effects in High School Majors
Gordon B. Dahl, Dan-Olof Rooth, and Anders Stenberg
NBER Working Paper No. 27618
July 2020, Revised June 2021
JEL No. I21,J24

ABSTRACT

This paper estimates peer effects both from parents to children and from older to younger siblings in academic fields of study in high school. Despite the importance of family peer effects, causal evidence is scarce due to correlated unobservables and a lack of data. Our setting is Sweden, where admission to oversubscribed majors is determined based on a student's GPA. Using a regression discontinuity design, we find large intergenerational and sibling peer effects that depend on the gender mix of a dyad. Younger brothers are 25% more likely to choose the major their older brother enrolls in, but 25% less likely to copy their older sister. Younger sisters copy older sisters (18% increase), but not older brothers. Effects vary based on birth spacing and whether a major is gender conforming. Turning to the effect of parental completion of a major, sons are 22% and 18% more likely follow in the footsteps of their fathers and mothers, respectively. In contrast, parents have little effect on their daughters' choices, except when a mother majors in the male-dominated program of Engineering. Since high school majors have a strong link to future occupation and earnings, these within family spillovers have long-term consequences for intergenerational income mobility and gender wage gaps. Finally, correlations greatly overstate these causal spillover effects, and miss heterogeneity by gender mix.

Gordon B. Dahl
Department of Economics
University of California, San Diego
9500 Gilman Drive #0508
La Jolla, CA 92093-0508
and NBER
gdahl@ucsd.edu

Dan-Olof Rooth
Swedish Institute for Social Research (SOFI)
Stockholm University
10691 Stockholm
Sweden
dan-olof.rooth@sofi.su.se

Anders Stenberg
Swedish Institute for Social Research (SOFI)
Stockholm University
10691 Stockholm
Sweden
anders.stenberg@sofi.su.se

1 Introduction

Parents and siblings are key players in an adolescent's network. From early childhood to the teenage years, family interactions occur daily and during a time when individuals are forming their own identities. Research in psychology has posited positive role model effects within families, where children look up to parents and older siblings and copy their choices. Families could also be a valuable vehicle for information transmission about different options. However, there are also theories which imply an opposite-signed recoil effect due to sibling rivalry, differentiation (a desire to be unique), or rebellion against parental expectations. The literature further hypothesizes that family influences will depend on the gender makeup of a sibling or parent-child pair (for overviews, see Bush and Peterson 2013; McHale et al. 2013; Raley and Bianchi 2006).

Family peer effects during the teenage years are particularly salient, as decisions are being made which will affect adult outcomes. One of the most important choices is what type of career path to pursue. In many countries throughout the world, including much of Europe, Latin America, and Asia, students choose a field specialization in secondary school (i.e., a "high school major").[1] This choice is made at the relatively young age of 15 or 16, when field preferences are still in flux and knowledge of different careers is limited, but before the natural distancing of siblings or full independence from parents. These high school major choices are consequential, as they play a foundational role in future labor market success. As we show in Dahl et al. (2020), the differential earnings returns across high school majors often rival the return to an additional two years of schooling. Traditionally male-dominated majors such as Engineering yield higher earnings, while female-dominated majors such as Humanities result in lower earnings. These effects are primarily driven by individuals with different high school majors ending up in different occupations, and to a lesser extent, different college majors.

In this paper, we examine the role of family peers – both parents and siblings – on high school major choices. Despite their importance, causal evidence on these types of spillovers remains scarce for two reasons. First, peer effects are notoriously difficult to estimate given the challenges identified by Manski (1993). Correlated unobservables are particularly likely

[1] Countries requiring students to choose fields in secondary school in Europe include the Czech Republic, Denmark, France, Italy, Norway, Poland, Spain, Sweden, and the United Kingdom; in Latin America include Argentina, Chile, Colombia, Cuba, Mexico, Paraguay, and Venezuela; and in Asia include Indonesia, Iran, Malaysia, Pakistan, the Philippines, and Saudi Arabia.

to create a bias when studying family spillovers.[2] High school major choices could be driven by common factors such as family income or shared environment, rather than a peer effect. Second, it is difficult to access and link datasets which contain the major choices of either siblings or of parents and their children. We overcome these challenges in the context of Sweden's secondary school system, where students choose between five academic majors: Engineering, Natural Science, Business, Social Science, and Humanities. Our setting is unique in that students were not allowed to choose which school to attend, so the choice concerns only majors and not institutions.

Academic majors are often oversubscribed, so admission is rationed based on a student's cumulative ninth grade GPA. Students rank their preferred majors and admission is determined using an allocation mechanism that is both Pareto efficient and strategy proof. Importantly, individuals around a major's GPA cutoff in a school region should be roughly similar on all observable and unobservable dimensions. This allows us to compare older siblings or parents just above versus just below major-specific admission cutoffs, and see (i) if their younger siblings or (ii) if their children are affected in their high school major choice. We use high-quality register data of applicants from 1977-1991 and merged child data from 2011-2019 to implement a regression discontinuity (RD) design.

We find that both older siblings and parents exert a significant influence on high school major choice. A younger sibling is 2.4 percentage points more likely to choose the same major as the one their older sibling enrolled in. Since younger siblings choose the same major 19.6% of the time, this translates to a 12% increase relative to the mean. Similar effect sizes are found for parental influences, with children being 11% more likely to choose the same major as their parent. However, these estimates mask a rich heterogeneity in spillover effects, as they do not account for the gender mix of siblings or parent-child pairs. As predicted by theories of family interactions, this turns out to be important, as spillover effects can be positive, negative, or zero for different gender-mix dyads.

Starting with younger brothers, they are 25% more likely to choose a particular major if their older brother enrolled in it. Examining birth spacing between the siblings provides further insight. For brother pairs, there is a relatively small and insignificant spillover effect if both brothers will be in school at the same time (i.e., within 3 years in age of each other).

[2] The other issues Manski identifies are endogenous group membership and reflection. The first is not an issue for family spillovers since individuals do not choose their siblings or children, and the second is not an issue if older siblings and parents influence younger siblings and children, but not the other way around.

In contrast, when brothers are more than 3 years apart, there is a substantial 42% increase in same major choice. This pattern is consistent with the older brother serving primarily as a role model or information provider when the age gap is large, but with sibling rivalry or differentiation counteracting such effects when the two brothers are close in age and will be in high school at the same time. These peer effects are large, with downstream effects for future occupational choices and earnings.

Younger brothers are also strongly influenced by older sisters, but in the opposite direction: a younger brother is 25% *less* likely to copy the major of an older sister. Birth spacing again turns out to be consistent with sibling rivalry or differentiation for contemporaneous school attendance. The estimated recoil effect rises to a 41% reduction for those within 3 years of age, while there is almost no effect for those further than 3 years apart. In other words, young brothers avoid their older sister's major, but only if they will overlap for a portion of their high school studies.

Our next finding is that younger sisters are heavily influenced by older sisters, with an 18% increase in the probability of choosing the same major. This effect is stronger for sisters closer in age, suggesting the opposite of sibling rivalry while they are in school together. Instead, sisters appear to play supportive roles for each other, consistent with a literature which argues that sisters are closer to each other compared to brothers or mixed-gender sibling pairs (Buist et al. 2002, McHale et al. 2013). Older brothers have little impact on sisters, with only a hint of an effect for those distant in age.

Turning to parental influences, we find the strongest effects for sons. When a father enrolls in a high school major, his son is 4.3 percentage points more likely to also choose it. If the father not only enrolls in a high school major, but also completes it, this effect rises to 5.6 percentage points.[3] This is a sizable 22% increase relative to the mean. Similarly, when a mother completes a high school major, her son is 18% more likely to choose it. In contrast, the effects for daughters are smaller and not statistically significant. These patterns are consistent with parents serving as role models and information providers, but more so for sons than daughters (Dryler 1998). As with the sibling estimates, these intergenerational effect sizes are large.

It is of interest to learn if observational data would yield similar findings. For years of

[3]For parents, we report effects related to both enrollment and completion. For siblings, we focus on enrollment, since younger siblings close in age make their choice before their older sibling completes a major.

schooling, Holmlund et al. (2011) report that the intergenerational correlation is more than twice the causal effect. We also find that observational data overstates spillover effects for high school majors, by a factor of 4.3 for siblings and 2.7 for parents and their children. Equally worrisome is that observational data misses out on much of the heterogeneity by gender mix, finding positive effects for all sibling and parent-child gender dyads. In contrast, our causal estimates find a negative recoil effect for older sisters on younger brothers, and no effect for older brothers on younger sisters. We also find little evidence of an effect for parents on daughters. The contrast between the correlational and causal estimates highlight the need to use a convincing research design to estimate family peer effects.

As a final exercise, we further break down these family spillovers based on the gender makeup of majors. Brothers copy brothers into the male-dominated and gender-neutral majors of Engineering, Natural Science, and Business, but not into the female-dominated majors of Social Studies and Humanities. We also find a stronger recoil effect for younger brothers when an older sister pursues a female-dominated major and that younger sisters are more likely to copy their older sisters into male-dominated and gender-neutral majors. The intergenerational results also exhibit some heterogeneity. The most interesting finding is that mothers who complete Engineering have a large effect on their children's choices: sons are 69% and daughters are 159% more likely to also choose Engineering.

The intergenerational and sibling results complement each other well. Because an individual affects both their children and their younger siblings, we are uniquely able to compare magnitudes. We find that an individual's major choice has similarly-sized influences on their children and their siblings.[4] When interpreting these results, it is useful to consider the time gaps in education decisions. The sibling spillovers occur relatively close in time, when curricula and information about a major are most similar, but usually before an older sibling has finished college or begun a job. In contrast, the intergenerational spillovers occur over 30 years later on average, but this gives the child an opportunity to observe a parent's occupation and earnings history. Both sets of results display a consistent pattern of the family as a mechanism for reinforcing gender stereotypical norms for males, but having either no impact or breaking gender stereotypical norms for females.

[4]It would also be interesting to compare whether an individual is more influenced by a sibling or a parent. This is unfortunately not possible with our data, as a series of reforms substantially reduced the number of oversubscribed majors after 1992 (see Section 2.2). We can still use major choices of those born after 1992 as outcomes in an intergenerational setting, but we do not have enough quasi-random variation due to oversubscription to identify sibling spillovers in majors for these later years.

In terms of family peer effects in education, our paper relates to three strands of research which span a variety of countries: family peer effects in years of schooling, course content, and choice of college.[5] To study the causal link between parent's and children's years of schooling, researchers have used school reforms, compulsory schooling laws, twins, and adoptions, often reaching different conclusions on the size of any intergenerational link (for reviews, see Black and Devereux 2011; Björklund and Jäntti 2009; Björklund and Salvanes 2011; Holmlund et al. 2011). In contrast, there are no similar causal studies for spillovers in siblings' years of schooling.

To study peer effects in course content, Joensen and Nielsen (2018) takes advantage of a pilot program in Denmark which lowered the cost of choosing advanced math and science classes. They find the younger siblings of those exposed to the program also chose more math and science courses. Peer effects are sizable and statistically significant for brother pairs, but not for other sibling-gender combinations.[6] Their paper is notable in providing some of the first causal evidence of sibling spillovers in course content. No studies have looked at intergenerational links in course content.

To study peer effects in college choices, a recent set of papers have used RD designs based on GPA cutoffs.[7] Altmejd et al. (forthcoming) looks at sibling spillovers in the U.S., Chile, Sweden, and Croatia. For the U.S., they find large effects of an older sibling barely getting into a college, both for whether the younger sibling attends any college or the same college. For the three other countries, there is compelling evidence that younger siblings follow their older siblings to the same college institution. But as Altmejd et al. write, "in contrast to the strong college-choice spillover effects, we find almost no influence on major choices" (p. 32).[8] For example, the reduced form estimate for Sweden of choosing the same major as an older sibling is estimated to be 0.000 (s.e.=0.002, see Table III, column 7). Similarly, Aguirre and

[5] There is also a literature which looks at sibling spillovers in school achievement. For two recent examples, see Nicoletti and Rabe (2019) and Quereshi (2018).

[6] The first stage for older sisters is on the border of statistical significance, which may partly explain the lack of a finding for older sister spillovers.

[7] Correlational estimates for sibling college choices also exist; for example, see Goodman et al. (2015).

[8] Prospective students in Chile, Sweden, and Croatia do not apply to colleges individually as in the U.S., but instead send in a single ranking of their college *plus* major preferences to a centralized administration that allocates students for the entire country. Students often prioritize majors over college institution, for example, by ranking Engineering at College A as their first choice and Engineering at College B as their second choice, rather than a different major being involved in their second choice. See Table III in the forthcoming *Quarterly Journal of Economics* version of Altmejd et al. (https://doi.org/10.1093/qje/qjab006), which finds strong evidence for target college+major and target college choice, but not for target major choice.

Matta (2018) and Dustan (2018) find evidence for sibling spillovers in institution choice in Chile and Mexico, but not for major. There are no similar studies for parent-child spillovers in college choice, likely due to the difficulty in obtaining the relevant intergenerational data.

More broadly, our paper is related to an emerging literature on other spillover effects within the family, both across siblings and across generations. A sampling of recent papers includes family spillovers in risky behavior (Altonji et al. 2017), military service (Bingley et al. 2021), entrepreneurship (Lindquist et al. 2015), cognitive and noncognitive skills (Lundborg et al. 2014), medical treatments and diagnoses (Daysal et. al. forthcoming; Fadlon and Nielsen 2019; Persson et al. 2021), labor supply (Nicoletti et al. 2018), disability and test scores (Black et al. 2021), and use of social insurance programs (Dahl et al. 2014).

Our paper makes two fundamental contributions. We are the first to causally estimate parent-to-child spillovers in the choice of major, finding large intergenerational effects. Second, we find strong evidence that the choice of what field to study in high school is influenced by siblings, in contrast to prior findings for college which only finds an effect for institution choice. In addition, our data allows us to explore heterogeneous effects across gender dyads, both for the intergenerational and sibling analyses. The pattern of results indicates the gender mix of dyads is of first order importance for both the size and direction of family peer effects. Moreover, there is a rich heterogeneity as a function of sibling spacing and the gender conformity of majors. Because majors have different earnings returns, these spillovers have long-term implications for intergenerational income mobility and gender wage gaps.

The remainder of the paper proceeds as follows. The next section describes Sweden's secondary education system and our unique data. Section 3 discusses identification using ranking lists in an RD design and Section 4 presents the first stage results. Section 5 presents our aggregate estimates of sibling and intergenerational peer effects, followed by analyses across gender dyads in Sections 6 and 7. Sections 8 and 9 present robustness checks and heterogeneous effects based on the gender conformity of a major. The final section concludes and discusses implications for policy.

2 Setting and Data

2.1 Admission to High School Majors in Sweden

After nine years of compulsory schooling, in the year individuals turn 16, students can apply to a field of study in secondary school (i.e., a high school major). In this paper, we focus on the five academic majors which are preparatory for university studies or direct entry into the labor market: Engineering, Natural Science, Business, Social Science, and Humanities. These academic majors take three years to complete, with Engineering having an optional fourth year.

The curricula for these majors differ substantially. There are a total of 96 credit hours spread over three years for the academic majors. The Engineering and Natural Science majors require between 32 to 37.5 credit hours in math and science courses, while Business, Social Science, and Humanities only require 14, 20, and 12 hours, respectively. Engineering additionally has 22.5 credit hours of technology classes, while Natural Science instead includes a more balanced set of social science, language, and art classes. Business requires 25 hours of specialized courses such as law and accounting, none of which are included in the other majors. Social Science requires up to 14.5 extra hours of social science classes compared to other majors, and Humanities requires up to 23 more credit hours of language classes. The conclusion is that the choice of a high school major is one of the most important decisions a teenager will make regarding their future career path.[9]

Individuals can also choose from a variety of non-academic majors which take two years to complete. These programs focus on vocational skills or general education, but not at the level required for admission to college. Since these non-academic programs are usually not oversubscribed, we cannot use our RD design for them, apart from including them as possible next-best options for older siblings and parents, and as possible choices for younger siblings and children. During the periods we study, roughly half of students enroll in an academic versus non-academic major.

If a major is oversubscribed, students compete for slots based on their application GPA. The application GPA is the average grade across 10-12 school subjects as of ninth grade. Grades in each subject range from 1 to 5 and have an approximate mean of 3 and standard deviation of 1 in the entire population of ninth grade students. Applicants have an extra 0.2 bonus points added to their GPA if they apply to a major which accepted 30% or less of their gender nationally in the prior year.

Students rank their preferences for up to 6 majors, and a central administration office

[9] College majors do not require specific high school majors for admission, but some college majors do require the completion of specific classes. For example, to major in engineering in college, one needs advanced math classes; these are included in the Engineering and Natural Science high school curricula. In Sweden, students can also take adult education courses after high school to satisfy the course requirements for college majors.

then allocates students. Admission decisions are made sequentially, with the highest-GPA applicant being admitted to their first-choice major, the second-highest GPA applicant being admitted to their highest-ranked major among the set of majors which still have space in them, and so forth. This "serial dictatorship" mechanism of allocating slots is both Pareto efficient and strategy proof, as long as 6 choices is not a binding constraint (Svensson 1999).[10]

The key factor which determines whether a major will be oversubscribed is the lumpiness of class sizes. Classes, and therefore majors, are often capped at 30 students. If there is only one class for a given major and 33 students list the major as their first choice, it will be impacted. In contrast, if only 27 students list it as their first choice, everyone will be admitted. Depending on expected demand for a major, there could be two or even three classes for a given major. This lumpiness often leads to a major being oversubscribed in a school region in one year, but not the next. It also means that the most popular majors are not necessarily the ones which will be oversubscribed.

After admission decisions are sent out in July, there can be reshufflings across different majors as students change their minds and new slots open up. These reallocations are not necessarily random. Luckily, we observe the actual admission decision, which is a binary function of the GPA cutoff. The sharp admission cutoff will be used in an RD design to instrument for either major enrollment or completion.

2.2 Data

We have information on the major ranking lists submitted by all students going back to 1977. This allows us to not only observe which major an applicant is admitted to, but also to account for next-best alternatives.[11] This is key for being able to identify an interpretable causal effect in the presence of an unordered choice set, as we explain in Section 3.1. Our data is similar to that in Dahl et al. (2020), which studies how different high school majors affect future earnings.

[10]Six choices does not appear to be a binding constraint, as only 0.2% of all applicants are admitted to their 6th choice and only 1.6% even list a sixth choice. Between 1982-84, bonus GPA points were also given for first and second choices on a ranking list, so students may not have revealed their true preferences. As we show, excluding these years does not materially affect our estimates.

[11]If an individual is admitted to either their first or second ranked choice, which happens 96% of the time, then we define these as the individual's preferred and next-best alternative majors, respectively. For individuals who are admitted to a third or lower ranked choice, their preferred choice is defined as the lowest-GPA choice above their accepted choice, and the next-best alternative as their accepted choice. For ease of exposition, in what follows we refer to the preferred major as the first-best choice, even if it wasn't the first choice on their list, and likewise the next-best alternative as the second-best choice.

Our sample period for parents and siblings is 1977-1991. During this period, the majors did not experience large changes and students were only choosing majors and not which school to attend. The vast majority of regions only had one school for each major, but in cases with more than one, students were not allowed to choose which to attend; instead, the assignment was made by school authorities based on geographical distance. Hence, during our time period there is only major choice and not school choice. Depending on the year, there are between 114 and 137 high school regions, with a median number of 432 applicants per year and school region. In 1992, Business, Social Science, and Humanities were merged into one major, only to re-emerge as separate majors again in 2011. The introduction of private schools and school choice, as well as other reforms, also substantially reduced the number of oversubscribed majors after 1992. These factors mean we can use the 1977-1991 period to study sibling spillovers, and 1977-1991 data for parents merged with 2011-2019 data for their children to study intergenerational spillovers. But we do not have enough quasi-random variation due to oversubscription to identify sibling spillovers during the 2011-2019 time period.[12] We use personal and family identification numbers to make links, and limit our analysis to full siblings (same biological mother and father) and biological children. In addition to linking major choices, we also add in demographic variables from population register data.

Figure A1 shows the distribution of major choices for applicants to an academic track for both of our time periods (panel A). Between 1977-1991, Engineering and Business were the most popular choices, with over one-fourth of applicants choosing each of these majors. Humanities was the least popular with fewer than 10% of individuals listing it as their first choice. The Engineering and Business shares decline substantially by 2011-2019, with Social Science seeing the largest increase.

A key difference across majors is the fraction of male versus female applicants. Panel B reveals that fewer than 20% of applicants who listed Engineering as their first-best choice were female. On the other end of the spectrum, in 1977-1991, Social Science was 70% female and Humanities was 85% female. In between are Natural Science and Business, which attract a roughly equal sex mix. This variation will allow us to explore whether male-dominated, gender-balanced, and female-dominated majors induce different types of family spillover patterns.

[12] For 2011-2019, we have roughly 6,000 observations where the older sibling applied to an oversubscribed program, which is far too small to be useful. In comparison, our 1977-1991 sibling sample has over 88,000 observations.

Our estimation samples for older siblings and parents are limited to those who have a first-best choice where demand exceeds supply. During 1977-1991, 55% of individuals applying to an academic program have a first-best choice which is oversubscribed. As Appendix Table A1 shows, the characteristics of students applying to oversubscribed programs is broadly similar to those applying to non-impacted programs. We drop all observations where GPA is missing or outside the range of 2.0 to 5.0 and limit our RD analysis to a sample window of -1.0 to +1.5 points around the normalized GPA cutoff. We also exclude a small number of school regions and years where two or more majors were combined. Finally, we cap family size at 5 siblings for the sibling analysis and 5 children for the intergenerational analysis, which drops 1.6% and 0.5% of the data, respectively.[13] This leaves us with an estimation sample of 88,164 sibling pairs and 168,933 parent-child pairs. There are 2.7 siblings on average in a family during 1977-1991, with older siblings having 1.1 younger siblings on average in our estimation sample. For the intergenerational sample, parents have 2.4 children on average, of which 1.4 are observed during 2011-2019 in our estimation sample.

2.3 Inferring GPA Cutoffs

Before continuing, we need to explain how GPA cutoffs are determined. While we observe the choice rankings for each individual and the admission decision, the GPA cutoff is not recorded in our dataset. Instead, we must infer the GPA cutoff from the data. Fortunately, the rules appear to have been followed, so this is relatively straightforward.

Each combination of year, school region, and major has the potential to be a competition for slots. We refer to these as "cells." Our RD design only applies to oversubscribed majors (i.e., competitive cells). If there are more applicants than slots, the admission GPA cutoff is inferred from the data. We limit our sample to cells where there is evidence for a sharp discontinuity, that is, where everybody above the GPA cutoff is admitted to the major and everybody below the cutoff is not.[14]

[13] In robustness checks, we further restrict the data to 2-sibling families (for the sibling analysis) and firstborns (for the intergenerational analysis).

[14] We allow for a small amount of noise in the data due to measurement error, which is likely during this time period since most variables were transcribed and entered by hand. For example, if one observation with a GPA of 3.8 is recorded as not admitted while all of the remaining observations higher than 3.3 are recorded as admitted, it is likely that either GPA or major was erroneously recorded. Our rule is to retain the cell if the "miscoded" observations represent less than ten percent of the observations at the given side of the cutoff. If the condition is met, we retain the cell, but drop the "miscoded" observations. This procedure drops just 0.3 percent of the data. We also require at least 25 applicants in a cell and at least 3 observations to the left of the cutoff.

One complexity is that there can be a mix of accepted and non-accepted individuals at a cutoff GPA. For example, if the cutoff is 3.2 in a cell, there may only be slots for 3 out of the 5 applicants with a GPA of 3.2. Ties can happen since GPA is only recorded to the first decimal. In this case, it is important to know how people at the cutoff with the same GPA were admitted. We found some documentation which indicated admission was random, but also documentation which said that sometimes secondary criteria such as math grades were used to break ties. Since we do not know the criteria used to break ties, we discard the observations at the cutoff GPA. This should not create a problem, as we are still able to identify a sharp discontinuity above and below this mixed-cutoff GPA. Continuing with the example of a mixed cutoff at 3.2, we would drop all individuals with a GPA exactly equal to 3.2 in the cell, but define the cutoff as 3.2 for the remaining observations in the cell.

When there is not a mix of accepted and non-accepted individuals at a cutoff, we simply define the cutoff GPA as the average between the two adjacent GPAs. So for example, if everyone with a GPA of 3.3 is not admitted and everyone with a GPA of 3.4 is admitted, we define the GPA cutoff for the cell as 3.35. To enable pooling of data across regions and years, we normalize the cutoff GPA to 0 in our RD regressions.

The distribution of cutoff GPAs is plotted in Appendix Figure A2. The figure also overlays the distribution of individual GPAs for applicants to oversubscribed academic majors. The mean cutoff GPA for oversubscribed majors is 3.4, a value which corresponds to the 18th percentile of GPAs among applicants to oversubscribed academic majors. To put this in further perspective, the mean cutoff GPA also corresponds to roughly the 60th percentile of GPAs in the sample of all ninth graders, including those who apply to nonacademic majors or do not continue on to high school. While the cutoffs vary from year to year, they are generally only binding for applicants with GPAs in the bottom half of our estimation sample.

In our setting, it is important not to confuse "oversubscription" with "highly competitive." This is because the cutoff for a major is determined by local supply versus demand, which varies from year to year within a school region. There is not a universal ordering of which majors are more likely to have higher admission cutoffs. For example, Engineering has a higher cutoff than Natural Science in 37% of years within the same school region on average, while the reverse is true in 25% of years. In 38% of years both programs either have open enrollment, or less commonly, identical cutoffs. Similar patterns are found for the other major combinations as reported in Appendix Table A2. Moreover, average cutoffs (conditional on having a cutoff) are broadly similar across majors, differing by less than 0.2 GPA points. These facts regarding the major cutoffs are useful to to keep in mind when interpreting the estimates, which will capture local average treatment effects for applicants around the cutoffs.

3 Model

3.1 Using Preferred and Next-Best Choices in an RD Design

Our goal is to estimate family peer effects in high school majors. In this section, we talk about modeling the effect of parents on their children, but the same ideas apply to older and younger siblings. As pointed out by Kirkeboen et al. (2016), with multiple unordered alternatives, identification requires not only randomization into majors, but also an accounting of second-best choices. We define dummy variables a_{jk} for $j = 1, ..., J$ and $k = 1, ..., K$ which equal 1 if a parent's preferred choice is j and next-best choice is k. The reduced form RD equation can be written as:

$$y_{jk} = \sum_{jk} a_{jk} 1[x < c_j] g^l_{jk}(c_j - x) + \sum_{jk} a_{jk} 1[x > c_j] g^r_{jk}(x - c_j) + 1[x > c_j]\theta + \delta_j + \tau_k + w'\gamma + e_{jk} \quad (1)$$

where we have omitted the subscript identifying parent-child pairs for convenience. The outcome variable y_{jk} is a dummy for whether a child chooses the same first-choice major as a parent who had preferred major j and next-best alternative k. The running variable x is the parent's GPA, c_j is the cutoff GPA for admission to major j, g^l_{jk} and g^r_{jk} are unknown functions to the left and right of the cutoffs, w is a set of pre-determined controls (year fixed effects and school region fixed effects), and e_{jk} is an error term. The parameters δ_j and τ_k are fixed effects for first and second best choices, respectively.

The main coefficient of interest is θ, which captures whether a child is more likely to copy the choice of a parent (or whether a younger sibling copies their older sibling). We will estimate θ both at the aggregate level, as well as by the gender mix of the parent and the child (or sibling pair). We will also explore further heterogeneity in θ based on whether a major is gender conforming.

We combine data across years and school regions by normalizing each cutoff to be 0, and adjust the GPA running variable accordingly. Equation 1 estimates the sharp design for how parents being admitted versus not admitted to a major affect children's choices (i.e., the reduced form). To scale this effect, we estimate fuzzy RDs using either a parent's enrollment in or completion of a major in first stage regressions.

We try various parameterizations for the functions g^r_{jk} and g^l_{jk}. Without any restrictions, there are 30 functions to the left of the cutoff and 30 functions to the right. Our most parsimonious parameterization allows just 2 slopes: a common slope to the left and a common slope to the right. Another possibility is to impose common slopes to the right of the cutoff for each of the 5 preferred choices (regardless of the next-best choice), and common slopes to the left of the cutoff for each of the 7 next-best choices (regardless of the preferred choice).[15] This parameterization links the slopes to the major a parent was admitted to. We find that the 2-slope model can be too restrictive, but that the 12-slope and 60-slope models reach similar conclusions. For comparison, Altmejd et al. (forthcoming) uses a two-slope model.

3.2 Threats to Validity

To have a valid RD design, the running variable cannot be perfectly manipulated around the cutoff. There is little chance of such manipulation in our setting. One reason is that the required GPA to gain admission to a major is not known in advance, and varies from year to year as a function of the number of applicants. Thresholds differ 83% of the time for majors with a cutoff in successive years, as illustrated in Figure A3.

One way to test for manipulation is to check for balance in pre-determined characteristics around the admission cutoff. We do this in Figure A4 for all applicants to an oversubscribed major during 1977-1991. There are no noticeable jumps at the cutoff, a finding which is confirmed with formal tests.[16] Another common test for manipulation is based on jumps in the density of the running variable at the cutoff. However, a standard McCrary (2008) test or the newer test proposed by Cattaneo et al. (2018) are not applicable since the cutoff is based on an order statistic.[17]

To identify the causal spillover effects of enrollment or completion, we need exclusion, monotonicity, and irrelevance conditions to hold. These are extra conditions which are not required for the reduced form of admission. The exclusion restriction requires that crossing the admission threshold for a major only affects outcomes through enrollment (or completion). For enrollment, being admitted but not enrolling provides little information or experience

[15]There are 7 second best choices, because we allow vocational and non-academic general majors as second-best choices, but only academic majors as first-best choices.

[16]For the 8 estimated jumps, the largest t-statistic is 1.4 and six are below 1.

[17]In ongoing research, Cattaneo, Dahl, and Ma are working on a proof for there being a spurious density discontinuity and ways to modify a density test to account for this.

with a major, and so it seems likely that the condition holds both for siblings and parents. For completion, it is possible that parents take some specialized courses because they are admitted to a major, but then do not complete the major. However, given the large differences in curricula across majors, most switching occurs early on during the fall of the first year so there is limited scope for this channel. The monotonicity assumption requires that crossing an admission threshold does not make an individual less likely to complete that major, which seems plausible.

Finally, we require the irrelevance condition discussed in Kirkeboen et al. (2016). In our context, this condition means that if crossing the GPA threshold for admission to a major (e.g., Engineering) does not cause a parent or older sibling to complete that major, then it does not cause them to complete a different major (e.g., Humanities) either. While this condition seems plausible in our setting, it is in theory possible that it does not hold for *completion* of a major. In contrast, we note that the irrelevance condition holds by construction for *admission* to a major. This is because we have a sharp discontinuity for admissions, where everybody above the GPA cutoff is admitted to the major. It is only when we use program completion to scale our reduced form estimates using a fuzzy RD that this issue arises. Likewise, it seems likely to hold for *enrollment* in a major, as almost everyone who is admitted to their first choice major initially enrolls in the major (see Figures 2 and A5) with switching mostly happening after enrollment.

To further probe the validity of our design, we conduct a placebo test. Specifically, we test whether a younger sibling's acceptance into a major affects their older sibling's ex-ante choice. Since the older sibling makes their major choice before their younger sibling knows if they have been accepted, there should not be an effect. In Table 1, we conduct this placebo test using the same RD regression specification as our main estimates (a specification which will be explained in detail below). For each gendered sibling pair, the placebo estimates are close to zero and not statistically significant. Note that a similar exercise cannot be performed for parents and children, since children made their major choices during a time period when few majors were oversubscribed.

4 First Stages

We begin our presentation of empirical results by documenting the first stages for both parents and older siblings. The first stages capture how admission to a major affects either enrollment or completion. Admission exhibits a sharp discontinuity as a function of an individual's GPA, jumping from 0 to 1 at the relevant cutoff. In other words, every applicant whose GPA exceeds the cutoff is accepted to their first choice major, while every applicant below is not. This is illustrated in Figure 1. Hence, we have a sharp RD design for the question of how admission to a major affects the choices of younger siblings or of children.

To scale this reduced form effect, we use a fuzzy RD (e.g., IV) design. For siblings, the relevant first stage is how an older sibling's admission affects their enrollment in a major. For parents, there are two possible first stages: one for enrollment in a major and one for completion of a major. We do not use major completion as a first stage for siblings because the older sibling often has not completed high school before the younger sibling submits their ranking list of preferred majors.

Figure 2 illustrates the first stages for siblings, separated by gender mix. Consider the top-left graph, which plots the probability an older brother enrolls in his first-best choice in normalized GPA bins. Everyone to the right of the vertical line is (initially) admitted to their first choice major, while everyone to the left is not (initially) admitted. Some people switch to other majors, usually before the school year begins. This reshuffling opens up slots for other students and explains why some individuals to the left of the admission cutoff are able to enroll in their first-best choice. Note that the density of normalized GPA is such that there are relatively few observations to the left of the cutoff, so a small drop in the enrollment shares in a GPA bin to the right of the cutoff can explain a large increase in the enrollment shares to the left of the cutoff.[18] Switching majors is not necessary random which is why we need to instrument for enrollment using the admission decision. There is a roughly 60% jump in the probability of enrollment at the cutoff; similarly-sized jumps occur in the other panels for the other sibling gender combinations.

The first-stage graphs for parental enrollment look similar to those for siblings, and are therefore relegated to the appendix (Figure A5). This is not surprising, as both the older sibling and parent samples span the same time period of 1977-1991. Parents have an additional first stage for completion of a major. In Figure 3 we plot this alternative first stage. Consider the top-left graph, which plots the probability a father completes his first-best

[18]For example, suppose there is an excess of applicants to the Engineering program, two classes of size 30, and 5 applicants just to the left of the cutoff. If 2 of the initially accepted students switch out of Engineering, this will open up 2 slots. If 2 of of the 5 individuals just to the left of the cutoff take these slots, then the enrollment rate for this group will be 40%.

choice in normalized GPA bins. A nontrivial fraction of fathers admitted to their first-best major do not complete it. This is mostly due to individuals switching to other majors, usually in their first year of studies; few drop out of high school entirely (approximately 5%). There is a roughly 45% jump in the probability a father completes his first-best choice at the cutoff. Similar jumps are found in the other panels of Figure 3.

In Table 2 we report the corresponding first stage regression estimates. The first stage estimates do not vary much by the gender mix of a dyad. The first stage enrollment estimates for older siblings or parents range from .59 to .64, while the first stage completion estimates for parents range from .43 to .47. The estimates are all highly significant, indicating there will not be a problem due to weak instruments.

5 Aggregate Estimates

We start our presentation of results by lumping all sibling pairs together and all child-parent pairs together regardless of gender. The dependent variable in these aggregate regressions is a dummy for whether a younger sibling's preferred major (or child's preferred major), defined as the first choice on their ranking list of majors, matches their older sibling's (or parent's) first-best choice. The RD regressions use the 12 slope model, separate linear functions of the running variable on each side of the cutoff, a window of -1.0 to 1.5, triangular weights, fixed effects for year and school region, dummies for preferred major interacted with dyad gender mix, and dummies for the next-best alternative major as discussed in Section 3.1. The regression also includes the demographic variables listed in Appendix Table A1. Standard errors are clustered at the family level.

Panel A of Table 3 reports both reduced form and fuzzy RD estimates for siblings. The reduced form estimate of 1.5 percentage points captures the increased probability a younger sibling copies the same first-choice major as their older sibling if their older sibling was admitted. Compared to the mean, younger siblings copy older sibling 19.6% of the time on average,[19] so the reduced form estimate represents an 8% increase. The IV estimate (i.e., the fuzzy RD estimate) scales the reduced form by the first stage probability an older sibling enrolls in a program, and is 2.4 percentage points, for a 12% increase relative to the mean.

In panel B, we report the reduced form and fuzzy RD estimates for all parent-child pairs.

[19]We calculate the copying average using the sample where an older sibling's GPA is within plus or minus 0.2 GPA points of the threshold for admission to a major.

The magnitudes are remarkably similar to the sibling estimates in panel A. The reduced form indicates that children are 1.4 percentage points more likely to choose their parent's first-choice major if their parent's GPA is just above the admission threshold. As a reminder, there are two possible first stages for parents: enrollment in a major and completion of a major. Starting with the enrollment IV estimate, there is a 2.3 percentage point increase in children choosing the same major as the one their parent enrolled in. Since children copy their parent's major choice 21.4% of the time on average, this translates to an 11% increase.[20] This is of a similar magnitude to the 12% increase we found for siblings. The other possible IV estimate captures how parental completion of a major affects a child's choices. The effect size rises to a 3.1 percentage point effect, or a 14% increase.

Comparison to Correlational Estimates. How do these peer effects compare to correlational estimates? For siblings, we calculate the fraction of younger siblings who list a major as their first choice if it is the one their older sibling enrolled in minus the fraction who choose it when their older sibling did not enroll in it. Doing this for each of the 5 majors and then taking the average across majors (weighted by the number of older siblings choosing each of the majors), yields our correlational estimate. This correlational estimate can be compared to the RD enrollment estimate appearing in Table 3, as it uses the same outcome and independent variable. As reported in Table 4, for siblings, we find a 10.3 percentage point increase in younger siblings copying their older sibling's enrollment choice. This is 4.3 times larger than the IV estimate of 2.4 percentage points. Doing a similar exercise for parents and their children, we find a correlational estimate of 6.3 percentage points, which is 2.7 times larger than the IV estimate of 2.3 percentage points. The differences between the correlational and IV estimates are statistically different from each other at the 1% significance level.

The upshot of these comparisons is that the observational estimates vastly overstate the magnitude of within-family peer effects.[21] This could be because unobserved factors which are common in the family drive similar major choices, rather than similar major choices reflecting a peer effect. However, there are still sizable roles for sibling and parental influences – just not as large as raw correlations would suggest. This parallels what has been found

[20] We calculate the copying average using the sample where a parent's GPA is within plus or minus 0.2 GPA points of the threshold for admission to a major.

[21] The differences between the correlational and IV estimates are not primarily due to IV capturing a local average treatment effect. If we restrict the correlational samples to an older sibling's or parent's GPA being within plus or minus 0.2 GPA points of the threshold for admission to a major, the correlational estimates are 3.0 times larger for siblings and 2.1 times larger for parents and their children.

for observational versus causal estimates of the intergenerational link in years of schooling. Summarizing papers in the Scandinavian context, Holmlund et al. (2011) report that the observed correlation across generations in years of schooling is between 0.2 and 0.3, while similar causal effects are markedly smaller at approximately 0.1. A similar comparison for correlational versus causal links for siblings' years of schooling is not possible, as no causal estimates exist.

While these aggregate effects are interesting, theories of family spillovers in psychology hypothesize that the gender mix of dyads should play important roles (e.g., McHale et al. 2013). We explore these heterogeneous effects in the next two sections.

6 Sibling Peer Effects by Gender Mix

This section reports reduced form and IV estimates of sibling peer effects as a function of both gender mix and birth spacing. The regression model builds on the aggregate specification used above, where to gain precision, we stack the data for different gender-mix combinations into a single regression. The dependent variable in these regressions is a dummy for whether a younger sibling's preferred major, defined as the first choice on their ranking list of majors, matches their older sibling's first-best choice.

The first two columns of Table 5 report sibling peer effects by gender mix. The magnitudes and even the direction of spillovers depend heavily on the gender composition of a dyad. Start with brother pairs. The reduced form shows that if an older brother is admitted to his first-best choice, there is a 4.3 percentage point higher probability his younger brother will choose the same major. The corresponding IV estimate implies a younger brother is 6.3 percentage points more likely to choose the same first-choice major if his older brother enrolled in it. Relative to the mean, this translates to a 25% increase. Turning to sister pairs, there is also a strong same-sex sibling effect. An older sister's enrollment increases the chances her younger sister will choose the same major by 3.9 percentage points, or 18% relative to the mean. These are sizable same-sex sibling spillovers.

The results for opposite-gender sibling pairs stand in sharp contrast. Consider older sisters and their younger brothers. In these families, the brothers recoil from the majors of their older sisters. A younger brother is 2.9 percentage points *less* likely to choose a major his older sister enrolls in. This amounts to a 25% reduction compared to the mean. When interpreting this effect, it is important to keep in mind that younger brothers are less likely to choose similar majors as older sisters (11.7%) versus older brothers (25.3%) due to the gender makeup of majors. Hence, an equally sized percentage point reduction for older sisters will translate to a larger percent reduction relative to the mean. Turning to older brother - younger sister sibling pairs, there is a small and statistically insignificant peer effect.

Graphs of the reduced form effects by sibling gender mix can be found in Figure 4. These graphs plot the probability a younger sibling chooses the same first-choice major as their older sibling, by gender mix. While the estimates appearing in Table 5 use the more flexible 12 slope model (see Section 3.1), Figure 4 illustrates effects corresponding to a 2 slope model. As we will show in the robustness section, the 2 slope model yields broadly similar point estimates, even though the more flexible model is our preferred specification. Moreover, these RD graphs show raw data which do not include any controls. Both of these differences imply that the visual jumps at the cutoff do not line up exactly with the results presented in the table.

To understand these RD graphs, focus on the older brother - younger brother group. As an older brother's GPA increases, there is an increasing chance a younger brother will choose the same major. This indicates that younger siblings are more likely to copy older siblings who are better students. But more importantly, there is a discrete jump up in this probability of choosing the same major at the cutoff. There are likewise noticeable jumps which accord with the regression results for older sister - younger brother and older sister - younger sister pairs.

Turning to correlational estimates by gender mix (see Appendix Table A3 panel A), these paint a biased and misleading picture, just as we found when doing the corresponding comparison at the aggregate level. The same-gender (brother-brother and sister-sister) observational estimates are larger than the causal RD estimates by a factor of 3. Moreover, the correlations indicate positive effects of older sisters on younger brothers, while the IV estimates reveal a negative recoil effect, and the strong positive correlational estimates of older brothers on younger sisters is not present in the IV results. All of these differences are statistically significant at the 1% level.

Estimates by Birth Spacing. To gain more insight into what might be driving these gender-mix patterns, we next estimate effects as a function of birth spacing. High school takes three years to complete for an academic major, except for Engineering which can take three or an optional four years. In Sweden, the age cutoff for starting school as a child is January 1. Hence if siblings are born within three calendar years of each other, they will attend high school at the same time (unless the older sibling has skipped a grade or the younger sibling has repeated a grade). Therefore, we estimate separate regressions based on whether a sibling pair will be attending high school concurrently, proxied by a birth spacing of 3 years or less.

Start with brother pairs again. Research in psychology suggests that older brothers often serve as role models which younger brothers copy, but also that sibling rivalry causes younger brothers to try and differentiate themselves (e.g., McHale et al. 2013). Which of these two efffects dominates is an empirical question. Table 5 shows that the positive peer effects observed in the first two columns are largely driven by brothers who will not be attending school at the same time. There is a 10.7 percentage point spillover effect, or a 42% relative increase, for brothers spaced more than 3 years apart. In contrast, younger brothers with concurrent school attendance experience an effect one-fourth as large, indicating that rivalry dominates any role model effect. This difference in peer effect magnitudes is statistically significant (p-value for difference = 0.01). These results are consistent with that sibling rivalry is most pronounced for brothers close in age and attending school at the same time, whereas older brothers more distant in age could provide additional information about a major, including any downstream consequences which might take time to materialize (such as post-high school job or college opportunities).

Sister pairs also exhibit heterogeneity by birth spacing, but with the reverse pattern. The positive peer effect is roughly twice as large for sisters attending school concurrently, although the estimates are noisy enough that the difference is not statistically significant. This pattern aligns with survey evidence that sister pairs are more attached to each other and supportive of each other's goals (Buist et al. 2002). This is not to say that rivalry does not exist among sisters, but that other mechanisms dominate.

Now turn to mixed gender sibling pairs. The recoil effect, where a younger brother turns away from his older sister's first-choice major, is largest when the two siblings are close in age. The recoil effect translates to a large and statistically significant 41% reduction for brothers who will be attending school at the same time as their older sister, but is close to zero otherwise (p-value for difference = 0.13). For older brother - younger sister sibling pairs, there is no spillover effect for concurrent school attendance and a hint of a positive effect for siblings spaced further apart. These results suggest different dominant mechanisms for relationships of opposite gender siblings. Brothers are most likely to want to differentiate themselves from an older sister if they will both be attending school together. And the fact that older brothers have less effect on younger sisters during concurrent school enrollment is consistent with closely spaced sisters having stronger ties.

Taken together, these results provide insight into models of sibling rivalry and differentiation, as well as siblings as role models and information providers, in ways which are gender specific. Interestingly, older sisters' peer influence (either negative or positive) is strongest when siblings are close in age, while older brothers are more influential when siblings are further apart.

While the heterogeneity by gender mix is interesting in itself, it also highlights that looking at peer effects without separating by gender mix can lead to misleading conclusions. If we ignore gender mix and instead impose a common effect for sibling spillovers, we estimate a more modest aggregate 12% effect (Table 3). But this masks the fact that some peer effects are positive and some are negative, and hence cancel each other out in an aggregate specification.

7 Intergenerational Peer Effects by Gender Mix

We now shift focus to intergenerational peer effects by gender mix. The dependent variable in these regressions is a dummy for whether a child's preferred major, defined as the first choice on their ranking list of majors, matches their parent's first-best choice. We use the same RD regression specifications as we do for siblings.

A large literature hypothesizes that fathers and mothers have different types of influences on their children, and that these effects additionally depend on the gender of a child (e.g., Bush and Peterson 2013, Holmlund et al. 2011, Raley and Bianchi 2006). Hence, in Table 6 we present separate estimates broken down by the gender mix of an intergenerational dyad.

We find substantial heterogeneity. To begin, there are large spillovers from fathers to sons. The table reports a 2.9 percentage point increase in a son's probability of choosing a major if his father was admitted to it, a 4.3 percentage point increase if his father enrolled in it, and a 5.6 percentage point increase if his father completed it. The IV estimate based on completion translates to a 22% increase relative to the mean. Mothers also have a strong influence on their sons. The IV estimate for completion translates to an 18% increase. When comparing these effects, it is important to keep in mind that sons are more likely to choose similar majors as their fathers than their mothers due to the gender makeup of majors (25.4% versus 17.2%). Hence, an equally sized percentage point increase for mothers translates into a larger percent increase relative to the mean.

In contrast to the effects found for sons, we find little evidence of parental influence on daughters' choices. The estimates for father-daughter and mother-daughter pairs are both relatively small and not statistically significant. A test for whether sons and daughters are equally affected by their parents is rejected with a p-value of 0.02 using the IV-completion estimates. This is a joint F-test for whether the father-son coefficient equals the father-daughter coefficient and the mother-son coefficient equals the mother-daughter coefficient.

These estimates reinforce the lesson learned from the sibling analysis: estimating aggregate peer effects which do not distinguish by gender can lead to incomplete and misleading conclusions. In this case, the reason is that sons are strongly affected by their parent's high school major choices, but daughters less so.

RD graphs illustrating the reduced form effects are found in Figure 5.[22] There are noticeable jumps at the GPA admission thresholds for sons choosing the same first-choice major as their fathers and as their mothers. In contrast, there are no noticeable jumps for daughters for either parent gender. This pattern of results lines up with the estimates presented in Table 6.

While not the focus of our analysis, the slopes as a function of the running variable provide additional evidence that parents have less impact on daughters versus sons in their major choices. Starting with the first figure, the higher a father's GPA, the more likely a son is to copy his father's first-best major choice. A similar pattern emerges for mothers and sons. But the slopes are relatively flat for daughters copying either their mothers or their fathers major choices. In other words, the pattern of the slopes suggest that sons are heavily influenced by their parent's GPA when deciding whether to choose a similar major, but daughters are not.

Panel B of Appendix Table A3 reports corresponding correlational estimates. As we found when doing a similar comparison at the aggregate level, these correlations overstate peer effects – by a factor of more than 2 for sons and by a factor of more than 4 for daughters. The correlations suggest that daughters are as influenced by their parents choices as are

[22] As with the sibling RD graphs, these represent a 2 slope model (instead of the more flexible 12 slope model used in the tables) and plot binned, raw data (not accounting for other control variables). Hence, the jumps at the cutoffs do not line up exactly with the more flexible model estimates presented in Table 6.

sons, but the IV estimates reveal that this is not the case. All of the correlational versus RD differences are statistically significant at the 5% level. The contrast between the correlational and causal estimates again highlights the need to use a convincing research design to estimate family peer effects.

Estimates by Birth Order. In Appendix Table A4, we report estimates based on whether a child is the firstborn in a family. The rationale for this split is that firstborn children do not have an older sibling making a high school major choice before they do, and so perhaps a parent's influence will be stronger. We find some evidence for this expected pattern. The IV estimates for each of the parent-child gender pairs are larger for firstborn versus non-firstborn children. However, these differences are not statistically significant. The largest gap is for sons of fathers, with firstborns experiencing a 7.8 percentage point increase in choosing the same major as the one their father completed compared to a 3.3 percentage point effect for later born children (p-value for difference = 0.14).

8 Specification Checks

Before continuing, we present a variety of robustness checks in Table 7, both for the sibling and intergenerational analyses. For simplicity, we focus on the reduced form results; IV robustness checks yield similar conclusions.

We start by probing the parameterization of the regression model. Column 1 repeats our baseline estimates from earlier tables for ease of comparison. In column 2, we use second-order polynomials instead of linear trends. The reduced form estimates are similar in magnitude, but the standard errors increase by roughly 30%.[23] Column 3 reveals the results are robust to cutting the window size in half. In columns 4 and 5 we omit 1982-84 (see footnote 10) and exclude the demographic control variables. The results remain similar to baseline. Column 6 limits the sibling sample to dyads from two-sibling families (i.e., families with exactly two children), and finds similar results.

We next show estimates for both the more parsimonious 2-slope model and the more flexible 60-slope model. The results are similar to the baseline 12-slope model, except that for the 2-slope model the mother-son estimate is no longer statistically significant. In Figure A6,

[23] Although not shown, the IV estimates become slightly larger since the first stage estimates shrink with a second-order polynomial. In this sense, the linear estimates we report as our baseline IV can be viewed as conservative.

we illustrate that the 2-slope model is too restrictive for some major choices, which motivates why we make the 12-slope model our baseline. The graphs in the first column plot averages of the binned outcome variable for younger siblings (panel A) and children (panel B) against the running variable, allowing for separate slopes for each of the five first-best choice majors to the right of the cutoff and a common slope to the left of the cutoff. The second column shows similar plots, but allowing separate slopes for each of the seven second-best major choices to the left of the cutoff and a common slope to the right of the cutoff. The graphs show that while many major-specific slopes are similar, not all are, which is why we choose the 12-slope model as our baseline.[24] For example, the slope for Natural Science (the dashed green line) is steeper to the right of the cutoff compared to other majors in the parent-child graph in panel B. We emphasize that Figure A6 is for illustrative purposes only; we never mix the 2-slope and 12-slope models in estimation.

In Appendix Table A5, we perform a different set of robustness exercises, where we explore alternative measures for whether a younger sibling or child copies their family peer. Our baseline definition uses whether the first-choice major on a younger sibling's or child's ranking list matches their older sibling's or parent's preferred major. If we instead use whether the younger sibling or child included their older sibling's or parent's preferred major in *any order* on their ranking list, the results hardly change. For the sibling analysis, the results are also similar if we use whether the younger sibling was accepted to or enrolled in the same major as their older sibling's preferred major, and become slightly smaller if we use major completion. For the intergenerational analysis, the magnitudes drop by roughly 30-40% if we use whether a child's acceptance, enrollment, or completion of the same major is the same as their parent's preferred choice. These patterns hold for both the reduced form and IV specifications.

9 Estimates by Gender Makeup of Majors

As a final exercise, we estimate separate sibling and intergenerational peer effects by the share of females in a major. This split is motivated by hypotheses that gender-specific role-modeling matters, and in particular that mothers in STEM fields can have a strong influence on their children. We continue to estimate the effect of a younger sibling (child) choosing the same major as their older sibling (parent). But instead of imposing a common treatment effect, we

[24] The graphs also make clear that the intercepts for the various first-best and second-best choices differ, which we always allow for in all of our regressions.

estimate separate effects by major type. We define a male-dominated major as Engineering (less than 20% female), female-dominated majors as either Social Science or Humanities (both over two-thirds female), and gender-balanced majors as Natural Science or Business (see Figure A1).

The results using this split for siblings by gender dyads are found in Table 8. Begin with brother pairs. Brothers copy brothers for the heavily male-dominated major of Engineering (8.6 percentage point effect, a 28% increase) and also into the gender-neutral majors of Natural Science and Business (6.0 percentage point effect, a 25% increase). But there is no statistical evidence that younger brothers follow older brothers into the female-dominated majors of Social Science or Humanities. Sister pairs, in contrast, do not exhibit such gender-conforming patterns. Sisters copy sisters into gender-neutral majors, but not the female-dominated majors. There is a large, but statistically insignificant, effect for Engineering.

The patterns are equally interesting for opposite-gender siblings. Younger brothers are less willing to choose a female-dominated major if their older sister enrolled in it. There is a 4.2 percentage point drop in the likelihood a younger brother chooses Social Science or Humanities if this would copy his sister. This translates into a 56% drop relative the mean; the percent effect is large because a relatively small fraction of brothers choose a female-dominated major to begin with. In contrast, older brothers have no significant effect on their younger sisters, regardless of the share of females in a major. These patterns are consistent with younger brothers being influenced by whether a choice is gender conforming, while younger sisters are more immune to such forces.

In Table 9 we perform a similar exercise for parents and their children. For sons, we find strong peer effects from both fathers and mothers, but only if the major is not female dominated. Sons are 5.9 and 6.6 percentage points, respectively, more likely to choose their father's major if it was male dominated or gender neutral. These effects translate to 22% and 26% effects relative to their respective means. Likewise, if the mother's major was male dominated or gender neutral, the estimate for sons yield an increase of 18.0 and 4.1 percentage points, respectively. For the male-dominated major of Engineering, the effect is particularly large: it translates to a 69% increase. In contrast, for female-dominated majors there is no statistical evidence that sons copy parents.

Turning to daughters, there is not much evidence for spillovers from either fathers or mothers. The lone exception is that mothers serve as a role model for Engineering. Daughters are 11.0 percentage points more likely to choose Engineering. Since so few daughters choose Engineering to begin with (6.9%), this is a substantial 159% increase relative to the mean. This quasi-experimental evidence lines up with the correlational evidence found by Jacobs et al. (2017), which documents that mothers with engineering careers strongly influence the chances their children, and especially their daughters, plan to pursue Engineering.

10 Conclusion

Our findings provide novel evidence that family peer effects play an important role in high school major choices. These major choices are made at age 16, when preferences over future careers are still in flux. While made at a young age, these choices result in earnings differences across majors which rival the return to an additional year or two of schooling. Most of the differences can be attributed to differences in future occupation, and to a lesser extent, college major (Dahl et al. 2020). Given the strong link to adult occupation and earnings, these within family spillovers have long-term consequences for intergenerational income mobility and gender wage gaps.

The family spillover effects we find are large. Younger brothers are 25% more likely to choose the major their older brother enrolled in, but 25% less likely to copy their older sister. Younger sisters are 18% more likely to copy their older sisters, but do not copy their older brothers. An interesting question for future research is why there are sibling peer effects in high school major choice, but not in college. Parental effects are likewise sizable and differ across gender dyads. Sons are 22% and 18% more likely follow in the footsteps of the major their fathers and mothers completed, respectively. Parents have little effect on their daughters' choices, except when a mother chooses Engineering. Another interesting question for future research is why this discrepancy for sons and daughters exists.

Gendered pathways drive many of our findings, both in terms of the sex makeup of sibling and parent-child pairs, and in terms of whether a major is gender conforming. Our results confirm theories and predictions that sibling and parent-child dynamics are differentially influenced by gender mix. Interestingly, we find that an individual's major choice has at least as large an influence on their sibling as their child, despite the literature's larger focus on parental influences (Björklund and Jäntti 2012, McHale et al. 2012). The patterns we observe also highlight that families play a key role in the development of gender norms, a mechanism emphasized by Bertrand (2011) and Brenøe (2018).

It follows that the family spillovers we identify matter for education policy, as they will amplify any direct effects of reforms. One policy is to expand slots in the higher-paying majors of Engineering, Natural Science, and Business. But a gender-blind expansion will risk reinforcing traditional gender norms for boys through family spillovers. More targeted policies, such as nudging girls into STEM or Business majors through outreach campaigns, or making it easier for girls to gain admission to these higher-paying majors, would help decrease gender occupation segregation and gender earnings gaps in the long run via family spillovers. For example, several studies find that exposure to female role models can increase female enrollment in traditionally male fields (Bettinger and Long 2005, Breda et al. 2020, Buckles 2019, Carrell et al. 2010, and Porter and Serra 2020). According to our results, family peer effects would contribute to making these effects even larger.

References

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

Altmejd, A., A. Barrios-Fernández, M. Drlje, J. Goodman, M. Hurwitz, D. Kovac, C. Mulhern, C. Neilson, and J. Smith (forthcoming): "O brother, where start thou? Sibling spillovers on college and major choice in four countries," *Quarterly Journal of Economics (https://doi.org/10.1093/qje/qjab006).*

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

Bertrand, M. (2011): "New perspectives on gender," in *Handbook of Labor Economics*, Elsevier, vol. 4, 1543–1590.

Bettinger, E. P. and B. T. Long (2005): "Do faculty serve as role models? The impact of instructor gender on female students," *American Economic Review*, 95, 152–157.

Bingley, P., P. Lundborg, and S. V. Lyk-Jensen (2021): "Brothers in arms: Spillovers from a draft lottery," *Journal of Human Resources*, 56, 225–268.

Björklund, A. and M. Jäntti (2009): "Intergenerational income mobility and the role of family background," in *Oxford Handbook of Economic Inequality*, Oxford University Press Oxford, 491–521.

——— (2012): "How important is family background for labor-economic outcomes?" *Labour Economics*, 19, 465–474.

Björklund, A. and K. G. Salvanes (2011): "Education and family background: Mechanisms and policies," in *Handbook of the Economics of Education*, Elsevier, vol. 3, 201–247.

Black, S. E., S. Breining, D. N. Figlio, J. Guryan, K. Karbownik, H. S. Nielsen, J. Roth, and M. Simonsen (2021): "Sibling spillovers," *The Economic Journal*, 131, 101–128.

Black, S. E. and P. J. Devereux (2011): "Recent developments in intergenerational mobility," in *Handbook of Labor Economics*, North Holland, vol. 4, 1487–1541.

Breda, T., J. Grenet, M. Monnet, and C. Van Effenterre (2020): "Do female role models reduce the gender gap in science? Evidence from French high schools," IZA Discussion Paper No. 13163.

Brenøe, A. (2018): "Origins of gender norms: Sibling gender composition and women's choice of occupation and partner," IZA Discussion Paper No. 11692.

Buckles, K. (2019): "Fixing the leaky pipeline: Strategies for making economics work for women at every stage," *Journal of Economic Perspectives*, 33, 43–60.

Buist, K. L., M. Deković, W. Meeus, and M. A. Van Aken (2002): "Developmental patterns in adolescent attachment to mother, father and sibling," *Journal of Youth and Adolescence*, 31, 167–176.

Bush, K. R. and G. W. Peterson (2013): "Parent–child relationships in diverse contexts," in *Handbook of Marriage and the Family*, Springer, 275–302.

Carrell, S. E., M. E. Page, and J. E. West (2010): "Sex and science: How professor gender perpetuates the gender gap," *Quarterly Journal of Economics*, 125, 1101–1144.

Cattaneo, M. D., M. Jansson, and X. Ma (2018): "Manipulation testing based on density discontinuity," *The Stata Journal*, 18, 234–261.

Dahl, G. B., A. R. Kostøl, and M. Mogstad (2014): "Family welfare cultures," *Quarterly Journal of Economics*, 129, 1711–1752.

Dahl, G. B., D.-O. Rooth, and A. Stenberg (2020): "High School Majors, Comparative (Dis)Advantage, and Future Earnings," NBER Working Paper No. 27524.

Daysal, N. M., M. Simonsen, M. Trandafir, and S. Breining (forthcoming): "Spillover effects of early-life medical interventions," *Review of Economics and Statistics.*

Dryler, H. (1998): "Parental role models, gender and educational choice," *British Journal of Sociology*, 375–398.

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

Fadlon, I. and T. H. Nielsen (2019): "Family health behaviors," *American Economic Review*, 109, 3162–91.

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.

Holmlund, H., M. Lindahl, and E. Plug (2011): "The causal effect of parents' schooling on children's schooling: A comparison of estimation methods," *Journal of Economic Literature*, 49, 615–651.

Jacobs, J. A., S. Ahmad, and L. J. Sax (2017): "Planning a career in engineering: Parental effects on sons and daughters," *Social Sciences*, 6, 2.

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

Kirkeboen, L. J., E. Leuven, and M. Mogstad (2016): "Field of study, earnings, and self-selection," *Quarterly Journal of Economics*, 131, 1057–1111.

Lindquist, M. J., J. Sol, and M. Van Praag (2015): "Why do entrepreneurial parents have entrepreneurial children?" *Journal of Labor Economics*, 33, 269–296.

Lundborg, P., A. Nilsson, and D.-O. Rooth (2014): "Parental education and offspring outcomes: evidence from the Swedish compulsory School Reform," *American Economic Journal: Applied Economics*, 6, 253–78.

Manski, C. F. (1993): "Identification of endogenous social effects: The reflection problem," *Review of Economic Studies*, 60, 531–542.

McCrary, J. (2008): "Manipulation of the running variable in the regression discontinuity design: A density test," *Journal of Econometrics*, 142, 698–714.

McHale, S. M., K. A. Updegraff, and S. D. Whiteman (2012): "Sibling relationships and influences in childhood and adolescence," *Journal of Marriage and Family*, 74, 913–930.

——— (2013): "Sibling relationships," in *Handbook of Marriage and the Family*, Springer, 329–351.

Nicoletti, C. and B. Rabe (2019): "Sibling spillover effects in school achievement," *Journal of Applied Econometrics*, 34, 482–501.

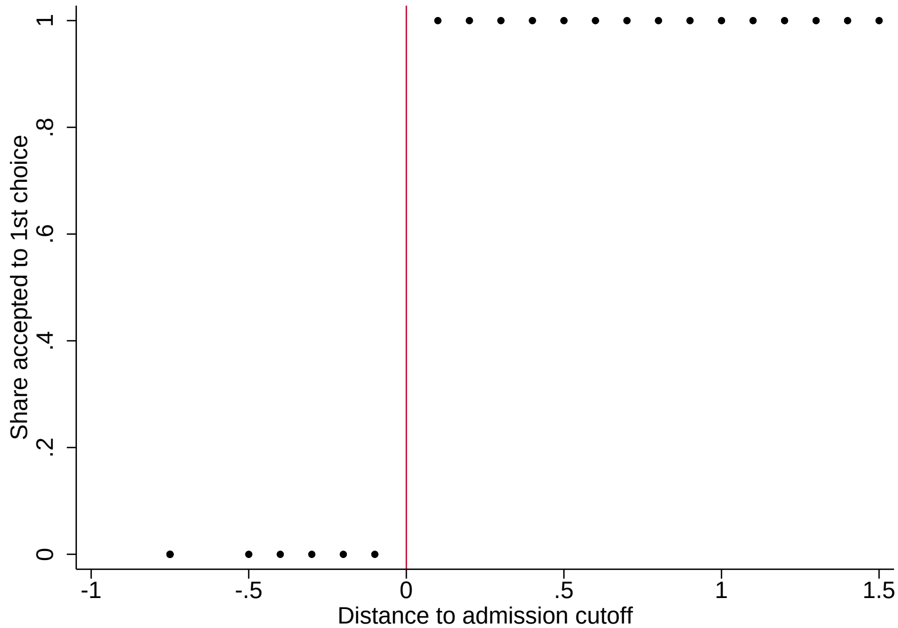
Nicoletti, C., K. G. Salvanes, and E. Tominey (2018): "The family peer effect on mothers' labor supply," *American Economic Journal: Applied Economics*, 10, 206–34.

Persson, P., X. Qiu, and M. Rossin-Slater (2021): "Family spillover effects of marginal diagnoses: The case of ADHD," NBER Working Paper 28334.

Porter, C. and D. Serra (2020): "Gender differences in the choice of major: The importance of female role models," *American Economic Journal: Applied Economics*, 12,

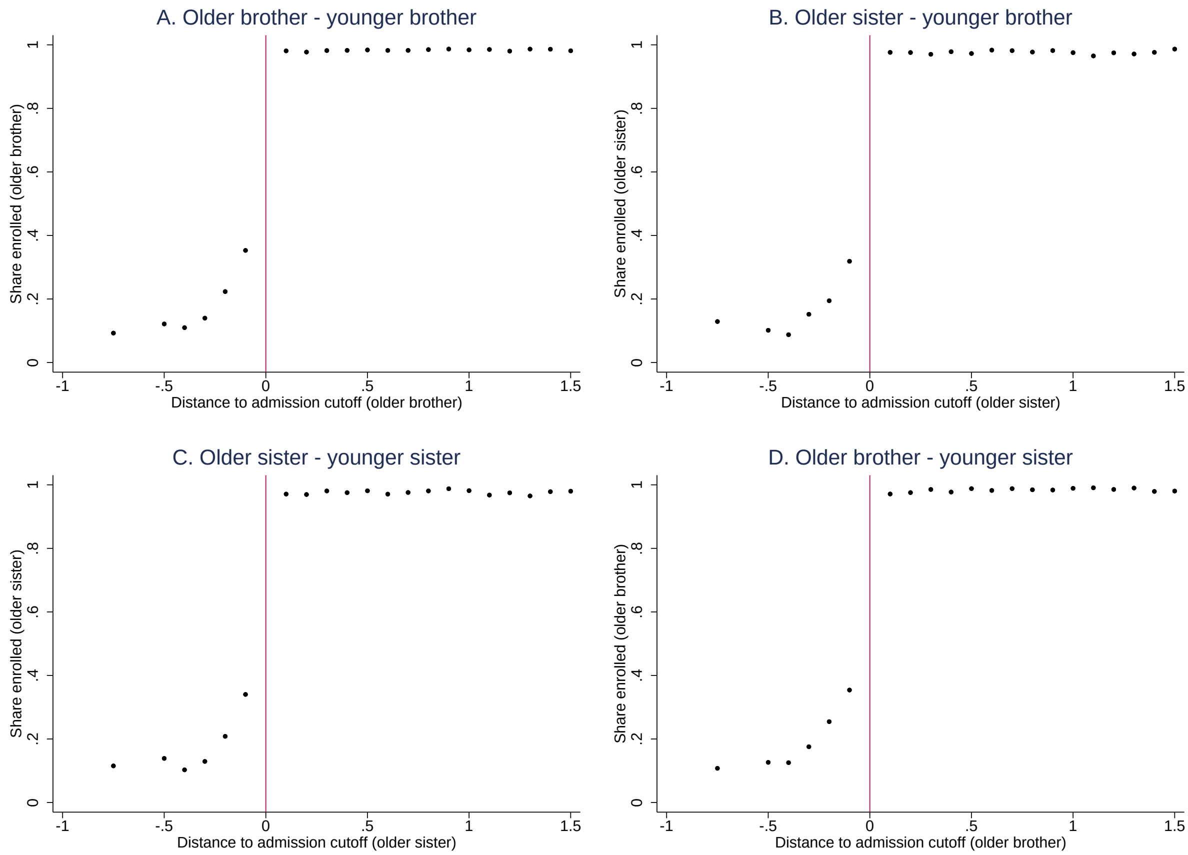
226–254.
QURESHI, J. A. (2018): "Siblings, teachers, and spillovers on academic achievement," *Journal of Human Resources*, 53, 272–297.
RALEY, S. AND S. BIANCHI (2006): "Sons, daughters, and family processes: Does gender of children matter?" *Annual Review of Sociology*, 32, 401–421.
SVENSSON, L.-G. (1999): "Strategy-proof allocation of indivisible goods," *Social Choice and Welfare*, 16, 557–567.

Figure 1. Share of individuals admitted to their first-choice major, 1977-1991.



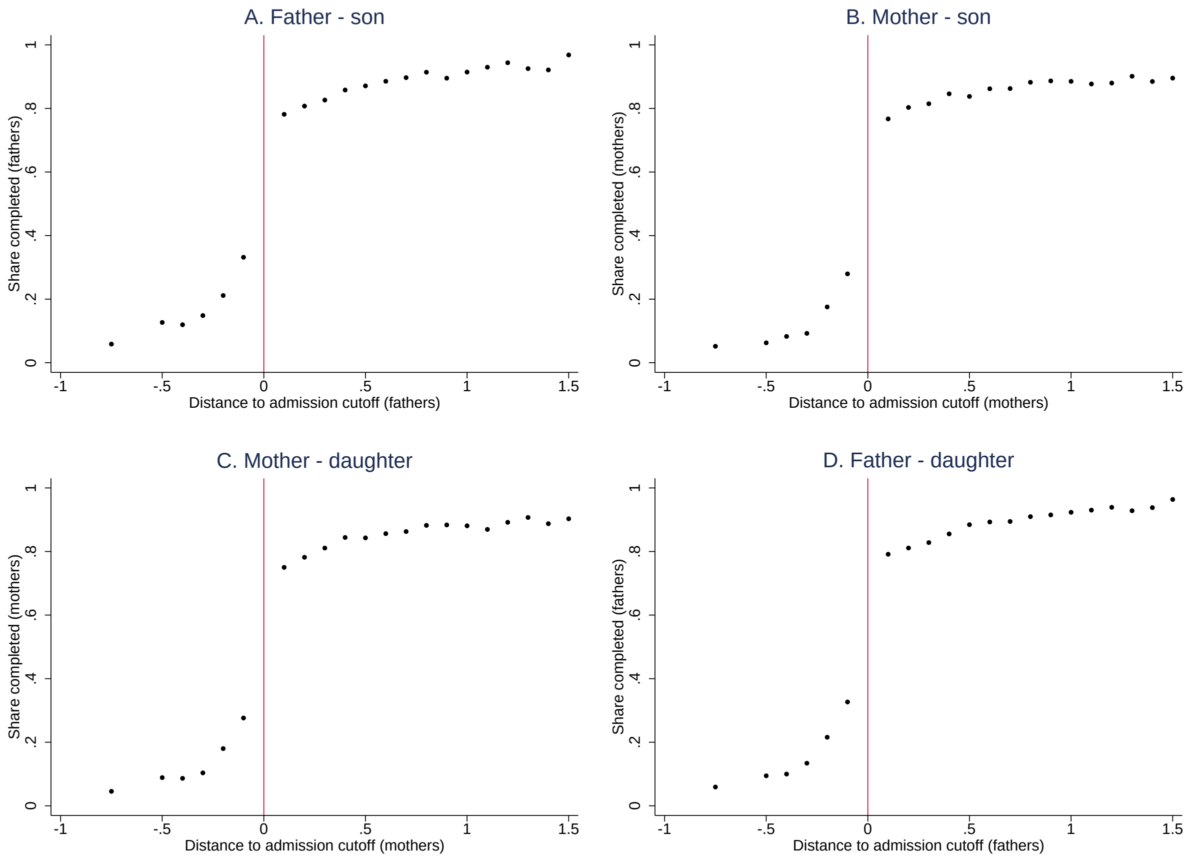
Notes: Each dot is the average share of individuals accepted to their first-best major choice as a function of their GPA in a 0.1 GPA bin, except for the leftmost dot which is a 0.5 bin due to sparsity. The vertical line denotes the admissions GPA cutoff (normalized to 0).

Figure 2. First stage: Share of older siblings enrolling in their first-choice major, by gender mix.



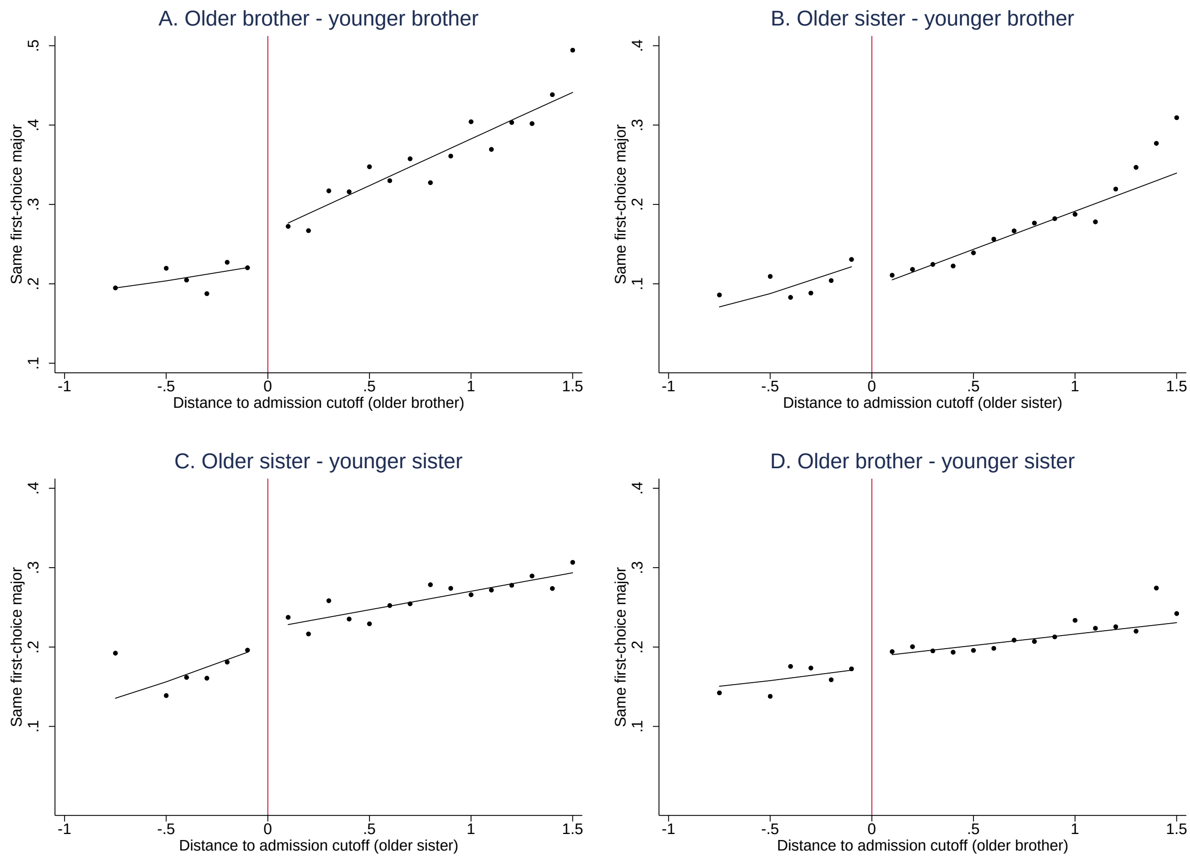
Notes: Each observation is the average share of older siblings who enroll in their first-best major choice as a function of their GPA. Each dot is a 0.1 GPA bin, except for the leftmost dot which is a 0.5 bin due to sparsity. The vertical lines denote the admissions GPA cutoff (normalized to 0). The number of observations in panel A is 22,841 (older brother - younger brother sample), in panel B 22,926 (older sister - younger brother sample), in panel C 21,020 (older sister - younger sister sample), and in panel D 21,377 (older brother - younger sister sample).

Figure 3. First stage: Share of parents completing their first-choice major, by gender mix.



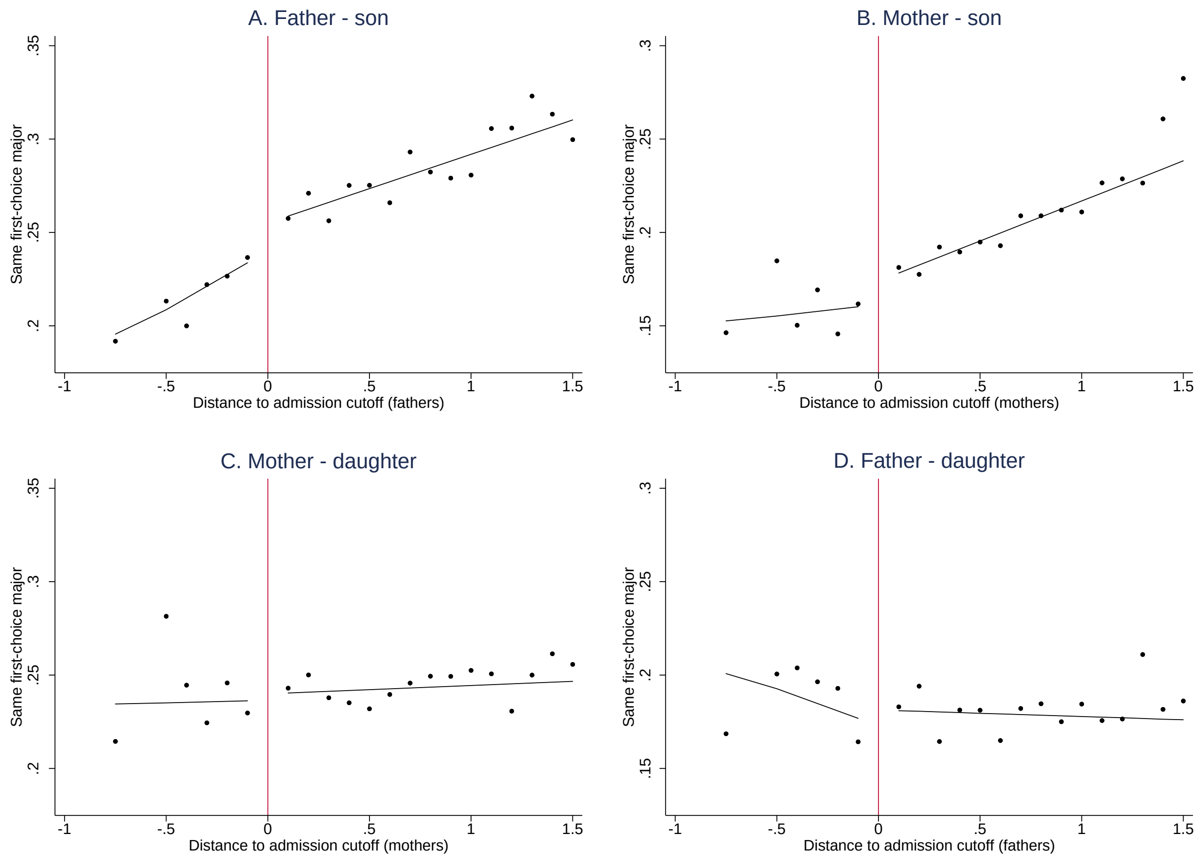
Notes: Each observation is the average share of parents who enroll in their first-best major choice as a function of their GPA. Each dot is a 0.1 GPA bin, except for the leftmost dot which is a 0.5 bin due to sparsity. The vertical lines denote the admissions GPA cutoff (normalized to 0). The number of observations in panel A is 37,390 (father-son sample), in panel B 49,057 (mother-son sample), in panel C 46,791 (mother-daughter sample), and in panel D 35,695 (father-daughter sample).

Figure 4. Reduced form: Probability a younger sibling chooses the same first-choice major as their older sibling, by gender mix.



Notes: Each observation is the average share of younger siblings whose first choice on their preference list matches the first-best major choice of their older sibling, as a function of their older sibling's GPA. Each dot is a 0.1 GPA bin, except for the leftmost dot which is a 0.5 bin due to sparsity. The vertical lines denote the admissions GPA cutoff for older siblings (normalized to 0). The estimated slopes are based on the 2-slope model, linear functions of GPA, a window of -1.0 to 1.5, and triangular weights. The number of observations in panel A is 22,841 (older brother - younger brother sample), in panel B 22,926 (older sister - younger brother sample), in panel C 21,020 (older sister - younger sister sample), and in panel D 21,377 (older brother - younger sister sample).

Figure 5. Reduced form: Probability a child chooses the same first-choice major as their parent, by gender mix.



Notes: Each observation is the average share of children whose first choice on their preference list matches the first-best major choice of their parent, as a function of their parent's GPA. Each dot is a 0.1 GPA bin, except for the leftmost dot which is a 0.5 bin due to sparsity. The vertical line denotes the admissions GPA cutoff for parents (normalized to 0). The slopes are based on the 2-slope model, linear functions of GPA, a window of -1.0 to 1.5, and triangular weights. The number of observations in panel A is 37,390 (father-son sample), in panel B 49,057 (mother-son sample), in panel C 46,791 (mother-daughter sample), and in panel D 35,695 (father-daughter sample).

Table 1. Placebo test: Does a younger sibling's first-choice major affect their older sibling's ex-ante choice?

	Reduced Form	IV-enrolled	Mean
(1) Impact on older brother			
Younger brother – older brother	.002	.003	[.324]
	(.010)	(.016)	
Younger sister – older brother	-.012	-.015	[.171]
	(.009)	(.014)	
(2) Impact on older sister			
Younger sister – older sister	.007	.010	[.232]
	(.011)	(.016)	
Younger brother – older sister	.004	.007	[.171]
	(.009)	(.015)	
N	91,683	91,683	

Notes: See notes to Table 5. The placebo regression estimates whether a younger sibling affects their older sibling, while Table 5 estimates whether an older sibling affects their younger sibling. Since the older sibling makes their major choice before their younger sibling, there should not be an effect in the placebo regression.
** p<0.10, ** p<0.05, *** p<0.01*

Table 2. First stage estimates for the probability of enrolling or completing a first-choice major.

Panel A: Siblings		**Panel B: Intergenerational**		
	Enrolled		Enrolled	Completed
(1) All		**(1) All**		
Older siblings	.618***	Parents	.604***	.449***
	(.008)		(.007)	(.007)
(2) By gender mix		**(2) By gender mix**		
Older brother – younger brother	.619***	Father – son	.602***	.432***
	(.011)		(.009)	(.009)
Older sister – younger brother	.636***	Mother – son	.609***	.465***
	(.012)		(.009)	(.009)
Older sister – younger sister	.624***	Mother – daughter	.612***	.458***
	(.013)		(.009)	(.009)
Older brother – younger sister	.602***	Father – daughter	.592***	.442***
	(.011)		(.009)	(.009)
N	88,164		168,933	168,933

Notes: The outcome variable is a dummy for whether the older sibling or parent enrolled (or completed) their first-best major, as a function of whether their GPA exceeded the admissions cutoff. Regressions use the 12-slope model, linear functions of GPA, a window of -1.0 to 1.5, triangular weights, fixed effects for year and school region, dummies for preferred major interacted with sibling/parent-child gender mix, dummies for next-best alternative major, and the demographic controls listed in Appendix Table A1. Standard errors in parentheses, clustered at the family level.
** p<0.10, ** p<0.05, *** p<0.01*

Table 3. Reduced form and IV estimates of the probability of choosing the same first-choice major as an older sibling or parent.

	Reduced form	IV-enrolled	IV-completed	Mean
Panel A: Siblings				
Older – younger sibling	.015**	.024**	--	[.196]
	(.007)	(.011)		
N	88,164	88,164	--	
Panel B: Intergenerational				
Parent – child	.014**	.023**	.031**	[.214]
	(.006)	(.009)	(.012)	
N	168,933	168,933	168,933	

Notes: The outcome variable is whether a younger sibling's or child's first choice on their preference list matches the first-best major choice of their older sibling or parent. IV-enrolled uses as a first stage whether the older sibling or parent enrolled in their first-best major, as a function of whether their GPA exceeded the admissions cutoff. IV-completed uses whether the parent completed their first-best major. Regressions use the 12-slope model, linear functions of GPA, a window of -1.0 to 1.5, triangular weights, fixed effects for year and school region, dummies for preferred major interacted with sibling/parent-child gender mix, dummies for next-best alternative major, and the demographic controls listed in Appendix Table A1. Dependent mean is calculated for the sample where an older sibling's or parent's GPA is within plus or minus 0.2 GPA points of the admission cutoff. Standard errors in parentheses, clustered at the family level.

** p<0.10, ** p<0.05, *** p<0.01*

Table 4. Correlational estimates.

	Correlational estimate	IV-enrolled	Difference
Panel A: Siblings			
Older – younger sibling	.103***	.024**	.079***
	(.002)	(.011)	(.011)
N	82,706	88,164	
Panel B: Intergenerational			
Parent – child	.063***	.023**	.040***
	(.001)	(.009)	(.009)
N	157,760	168,933	

Notes: Correlational estimates are based on the fraction of younger siblings/children who list a major as their first choice if it is the one their older sibling/parent enrolled in minus the fraction who choose it when their older sibling/parent did not enroll in it. This is done for each of the 5 majors and averaged across majors (with weights equal to the number of older siblings/parents choosing each of the majors). Bootstrap standard errors based on 1,000 replications. For IV-enrolled estimates, see notes to Table 3.

** p<0.10, ** p<0.05, *** p<0.01*

Table 5. Sibling estimates by gender mix and birth spacing.

	Any age gap		**Age gap ≤ 3 years (concurrent school enrollment)**		**Age gap > 3 years**		
	Reduced form	IV-enrolled	Reduced form	IV-enrolled	Reduced form	IV-enrolled	Mean
(1) Impact on younger brother							
Older brother – younger brother	.043***	.063***	.019	.027	.074***	.107***	[.253]
	(.010)	(.015)	(.013)	(.021)	(.015)	(.022)	
Older sister – younger brother	-.026***	-.029**	-.037***	-.048**	-.010	-.006	[.117]
	(.009)	(.014)	(.013)	(.019)	(.014)	(.020)	
(2) Impact on younger sister							
Older sister – younger sister	.025**	.039**	.036**	.049**	.013	.025	[.215]
	(.011)	(.016)	(.015)	(.021)	(.017)	(.025)	
Older brother – younger sister	.009	.017	-.001	-.001	.023	.039*	[.187]
	(.009)	(.014)	(.012)	(.019)	(.014)	(.021)	
N	88,164	88,164	52,270	52,270	35,894	35,894	

Notes: The outcome variable is whether a younger sibling's first choice on their preference list matches the first-best major choice of their older sibling. IV-enrolled uses as a first stage whether the older sibling enrolled in their first-best major, as a function of whether their GPA exceeded the admissions cutoff. Regressions use the 12-slope model, linear functions of GPA, a window of -1.0 to 1.5, triangular weights, fixed effects for year and school region, dummies for preferred major interacted with sibling gender mix, dummies for next-best alternative major, and the demographic controls listed in Appendix Table A1. Dependent mean is calculated for the sample where an older sibling's GPA is within plus or minus 0.2 GPA points of the admission cutoff. Standard errors in parentheses, clustered at the family level.

** p<0.10, ** p<0.05, *** p<0.01*

Table 6. Intergenerational estimates by gender mix.

	Reduced form	IV-enrolled	IV-completed	Mean
(1) Impact on sons				
Father – son	.029***	.043***	.056***	[.254]
	(.008)	(.012)	(.016)	
Mother – son	.015**	.024**	.031**	[.172]
	(.007)	(.011)	(.014)	
(2) Impact on daughters				
Mother – daughter	.004	.010	.015	[.244]
	(.008)	(.012)	(.015)	
Father – daughter	.007	.014	.021	[.185]
	(.008)	(.012)	(.015)	
N	168,933	168,933	168,933	

Notes: The outcome variable is whether a child's first choice on their preference list matches the first-best major choice of their parent. IV-enrolled uses as a first stage whether the parent enrolled in their first-best major, as a function of whether their GPA exceeded the admissions cutoff. IV-completed instead uses whether the parent completed their first-best major. Regressions use the 12-slope model, linear functions of GPA, a window of -1.0 to 1.5, triangular weights, fixed effects for year and school region, dummies for preferred major interacted with parent-child gender mix, dummies for next-best alternative major, and the demographic controls listed in Appendix Table A1. Dependent mean is calculated for the sample where a parent's GPA is within plus or minus 0.2 GPA points of the admission cutoff. Standard errors in parentheses, clustered at the family level.

** p<0.10, ** p<0.05, *** p<0.01*

Table 7. Sibling and intergenerational reduced form estimates using alternative specifications.

	Baseline	Quadratic	Smaller bandwidth	Excluding 1982-84	No demo. controls	2-sibling families	2-slope model	60-slope model
(1) Impact on younger brother								
Older brother – younger brother	.043***	.044***	.041***	.046***	.039***	.050***	.044***	.043***
	(.010)	(.013)	(.012)	(.012)	(.010)	(.015)	(.010)	(.010)
Older sister – younger brother	-.026***	-.024*	-.020*	-.021*	-.028***	-.028***	-.031***	-.024**
	(.009)	(.012)	(.011)	(.012)	(.009)	(.014)	(.009)	(.010)
(2) Impact on younger sister								
Older sister – younger sister	.025**	.027*	.027**	.037***	.024**	.031*	.019*	.026**
	(.011)	(.014)	(.013)	(.014)	(.011)	(.017)	(.011)	(.012)
Older brother – younger sister	.009	.011	.009	.015	.006	-.003	.010	.010
	(.009)	(.012)	(.011)	(.011)	(.009)	(.014)	(.009)	(.010)
N	88,164	88,164	62,205	62,375	88,164	42,182	88,164	88,164
(3) Impact on sons								
Father – son	.029***	.036***	.030***	.041***	.028***		.030***	.028***
	(.008)	(.010)	(.009)	(.009)	(.008)		(.008)	(.008)
Mother – son	.015**	.022**	.021**	.021**	.014*		.009	.016**
	(.007)	(.010)	(.008)	(.009)	(.007)		(.007)	(.008)
(4) Impact on daughters								
Mother – daughter	.004	.011	.009	.006	.003		-.002	.005
	(.008)	(.010)	(.009)	(.009)	(.008)		(.008)	(.008)
Father – daughter	.007	.014	.011	.012	.006		.008	.006
	(.008)	(.010)	(.009)	(.009)	(.008)		(.008)	(.008)
N	168,933	168,933	122,260	122,466	168,933		168,933	168,933

Notes: See notes to Tables 5 and 6. Column 2 includes quadratic functions of the running variable on both sides of the cutoff and column 3 reduces the bandwidth in half. Column 4 excludes the years 1982-84, when bonus GPA points were added for the first and second choices on an individual's ranking list, and column 5 excludes the demographic controls listed in Table A1. Column 6 only includes families with two siblings. Columns 7 and 8 use the 2-slope and 60-slope models. Standard errors in parentheses, clustered at the family level.

** p<0.10, ** p<0.05, *** p<0.01.*

Table 8. Sibling estimates as a function of the gender makeup of majors.

	Reduced form	IV-enrolled	Mean
(1) Impact on younger brother			
Older brother – younger brother:			
Male-dominated major (E)	.058***	.086***	[.306]
	(.016)	(.023)	
Gender-neutral major (N+B)	.042***	.060***	[.242]
	(.014)	(.020)	
Female-dominated major (S+H)	.010	.016	[.139]
	(.017)	(.023)	
Older sister – younger brother:			
Male-dominated major (E)	.019	.030	[.296]
	(.067)	(.094)	
Gender-neutral major (N+B)	-.020	-.023	[.146]
	(.014)	(.020)	
Female-dominated major (S+H)	-.035***	-.042***	[.075]
	(.011)	(.015)	
(2) Impact on younger sister			
Older sister – younger sister:			
Male-dominated major (E)	.057	.099	[.138]
	(.047)	(.078)	
Gender-neutral major (N+B)	.041**	.059**	[.242]
	(.017)	(.023)	
Female-dominated major (S+H)	.009	.015	[.195]
	(.015)	(.021)	
Older brother – younger sister:			
Male-dominated major (E)	-.011	-.011	[.090]
	(.011)	(.017)	
Gender-neutral major (N+B)	.024	.036	[.269]
	(.016)	(.022)	
Female-dominated major (S+H)	.024	.036	[.252]
	(.021)	(.029)	
N	88,164	88,164	

Notes: See notes to Table 5. The regressions differ by allowing for heterogeneous effects based on the gender makeup of majors. Majors are denoted in parentheses by their first letters: E, N, B, S, H stand for Engineering, Natural Science, Business, Social Science, and Humanities, respectively. Standard errors in parentheses, clustered at the family level.

** p<0.10, ** p<0.05, *** p<0.01*

Table 9. Intergenerational estimates as a function of the gender makeup of majors.

	Reduced form	IV-enrolled	IV-completed	Mean
(1) Impact on sons				
Father – son:				
Male-dominated major (E)	.023*	.039**	.059**	[.268]
	(.012)	(.019)	(.026)	
Gender-neutral major (N+B)	.039***	.056***	.066***	[.250]
	(.011)	(.016)	(.019)	
Female-dominated major (S+H)	.015	.024	.035	[.229]
	(.016)	(.021)	(.029)	
Mother – son:				
Male-dominated major (E)	.079*	.166*	.180**	[.262]
	(.043)	(.094)	(.089)	
Gender-neutral major (N+B)	.023**	.034**	.041**	[.194]
	(.010)	(.014)	(.017)	
Female-dominated major (S+H)	.004	.009	.014	[.144]
	(.010)	(.014)	(.019)	
(2) Impact on daughters				
Mother – daughter:				
Male-dominated major (E)	.043*	.094*	.110**	[.069]
	(.025)	(.054)	(.053)	
Gender-neutral major (N+B)	.001	.005	.009	[.220]
	(.011)	(.015)	(.018)	
Female-dominated major (S+H)	.005	.011	.017	[.278]
	(.011)	(.016)	(.022)	
Father – daughter:				
Male-dominated major (E)	.005	.012	.027	[.065]
	(.008)	(.013)	(.018)	
Gender-neutral major (N+B)	.002	.007	.011	[.239]
	(.012)	(.017)	(.019)	
Female-dominated major (S+H)	.023	.035	.047	[.326]
	(.018)	(.024)	(.031)	
N	168,933	168,933	168,933	

Notes: See notes to Table 6. The regressions differ by allowing for heterogeneous effects based on the gender makeup of majors. Majors are denoted in parentheses by their first letters: E, N, B, S, H stand for Engineering, Natural Science, Business, Social Science, and Humanities, respectively. Standard errors in parentheses, clustered at the family level.

** p<0.10, ** p<0.05, *** p<0.01*

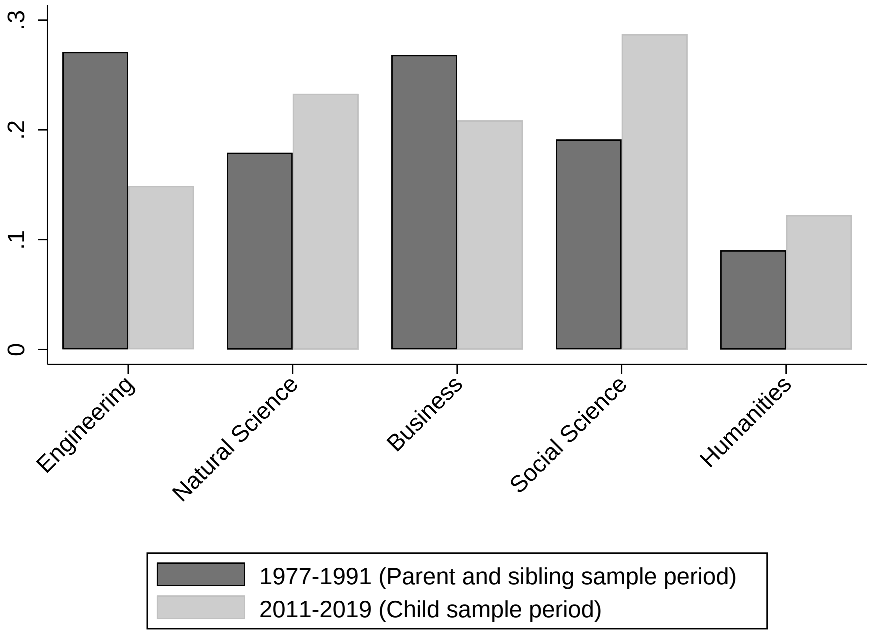
Appendix Figures and Tables
for Online Publication

"Intergenerational and Sibling Peer Effects in High School Majors"

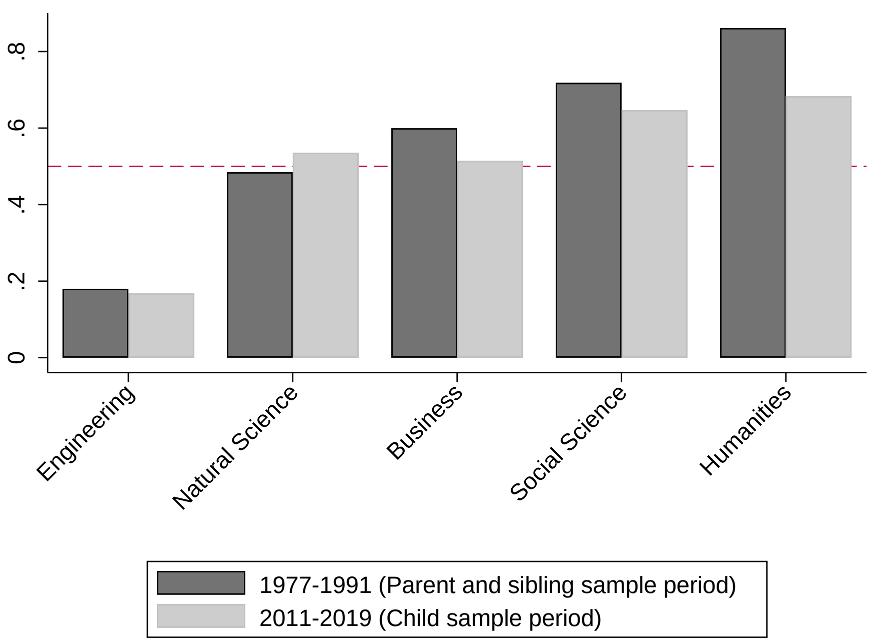
Gordon B. Dahl, Dan-Olof Rooth, and Anders Stenberg

Figure A1. Applications to academic high school majors, 1977-1991 and 2011-2019.

Panel A: Share of applicants by major

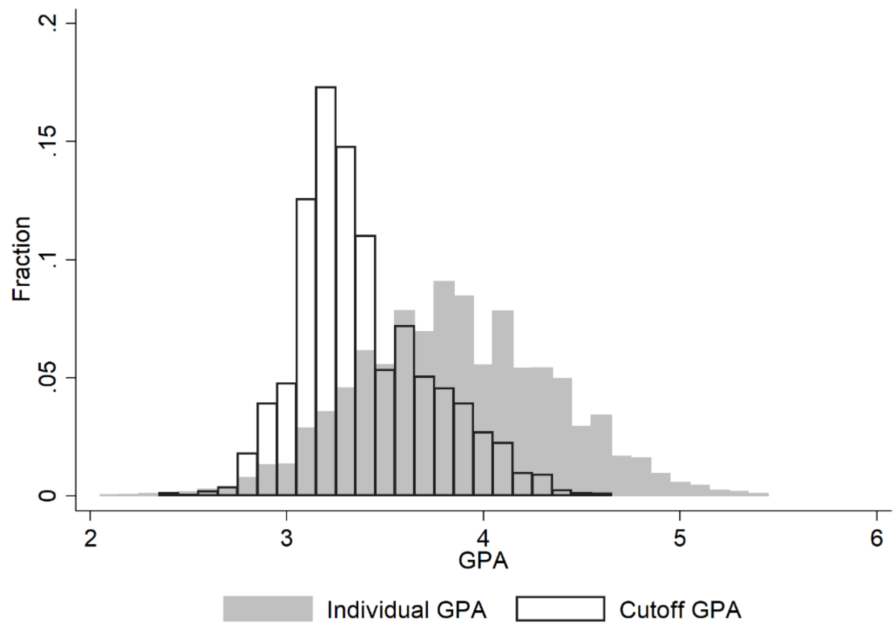


Panel B: Female share of applicants



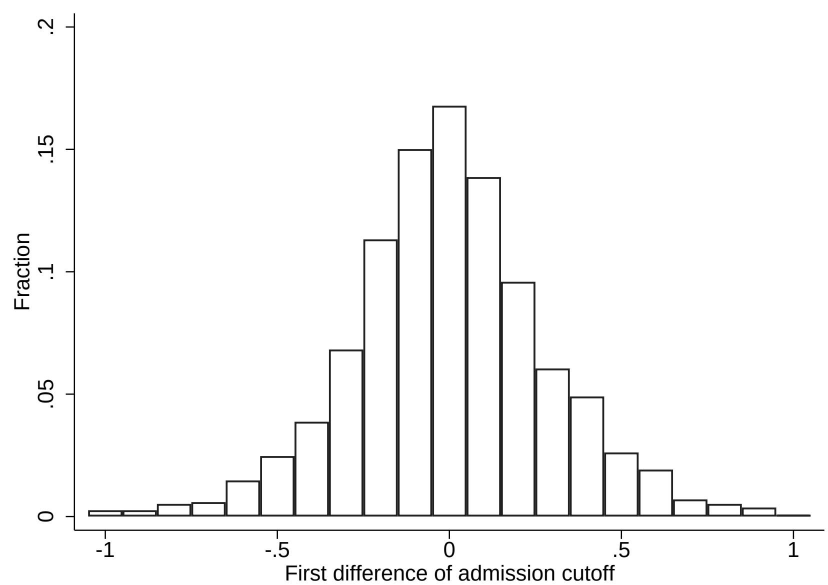
Notes: All applications to academic majors. For the years 1977-1991, N=607,767. For 2011-2019, N=558,442. The share in Humanities 2011-2019 also includes those in Arts. The dashed line marks a balanced gender composition.

Figure A2. Distribution of individual GPAs and major cutoff GPAs.



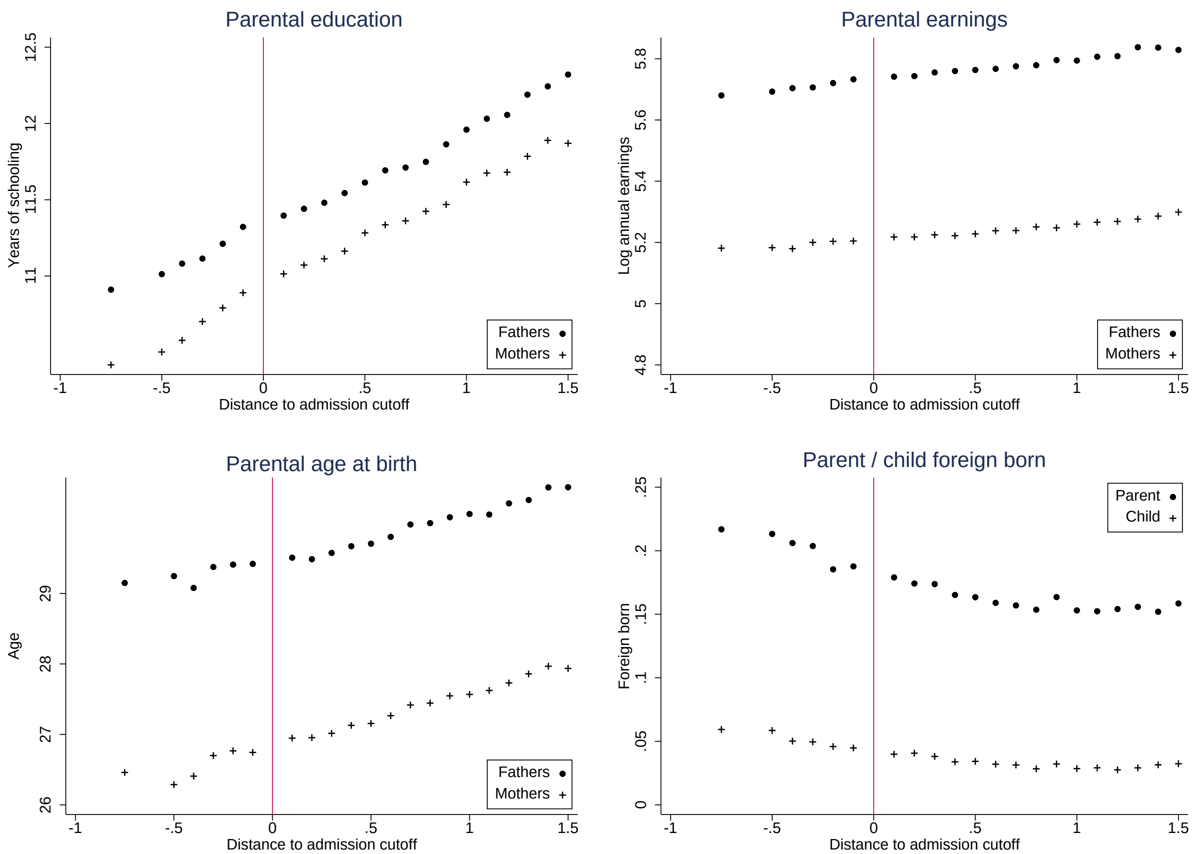
Notes: The white bars plot the distribution of cutoff GPAs for competitive programs, which vary by major, year, and school region. There are 3,487 competitive programs in our estimation sample. The grey bars plot the distribution of GPA for individuals in oversubscribed programs (N=263,878).

Figure A3. Distribution of current minus lagged admission cutoff GPA, 1977-1991.



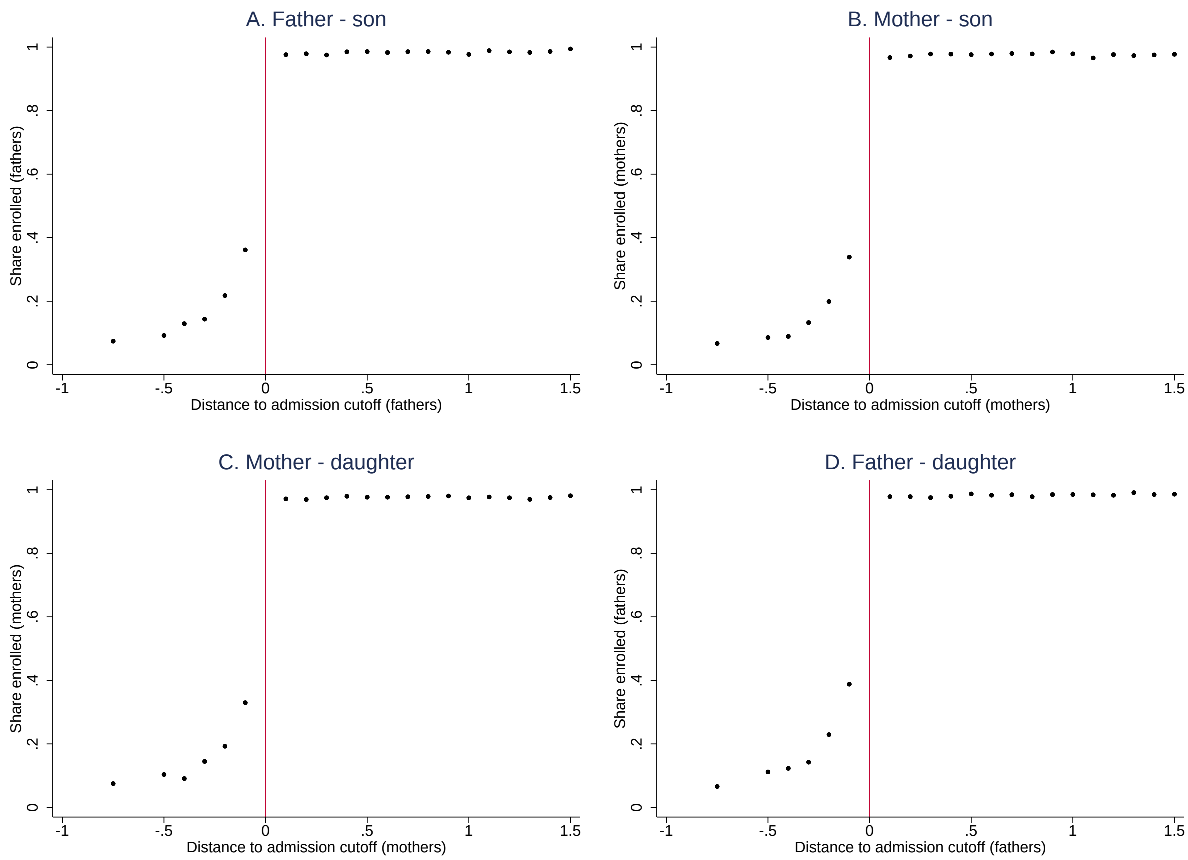
Note: Sample limited to academic majors which are oversubscribed two years in a row in a school region.

Figure A4. Smoothness of predetermined demographic variables at the cutoff.



Notes: Each marker is the average for the relevant outcome in a 0.1 GPA bin, except for the leftmost marker which is a 0.5 bin due to sparsity. The vertical lines denote the admissions GPA cutoff for individuals in oversubscribed programs between 1977-1991, (normalized to 0). Parent foreign born is a dummy for whether at least one parent is foreign born. Parents here refer to the parents of applicants during 1977-1991 (i.e., these are the grandparents of the children in our intergenerational sample).

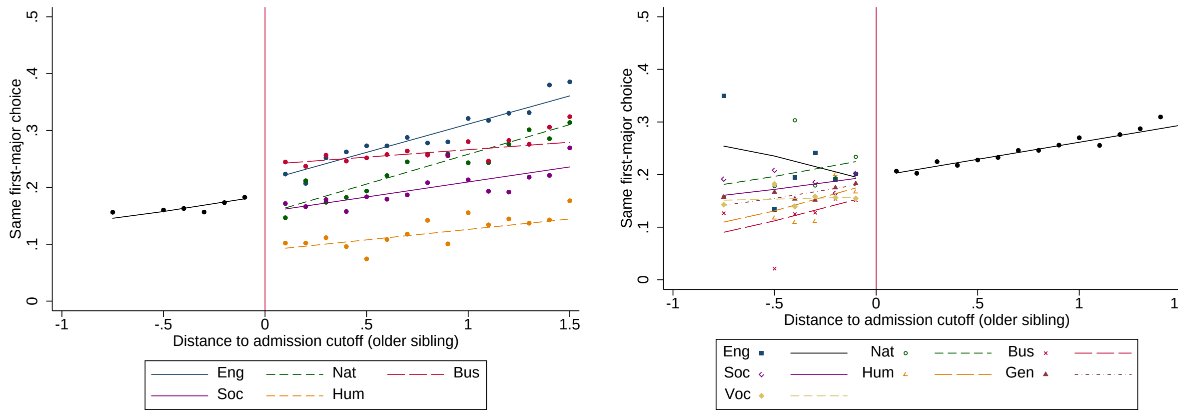
Figure A5. First stage: Share of parents enrolling in their first-choice major, by gender mix.



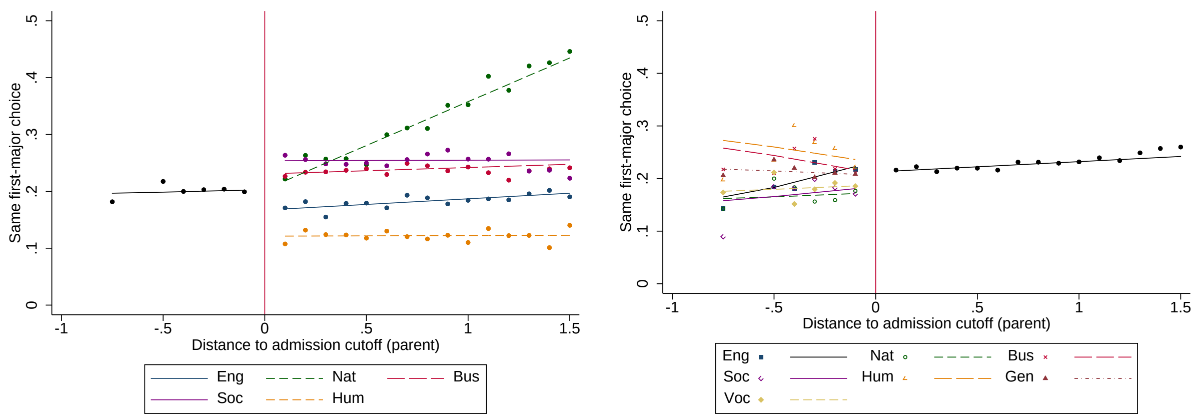
Notes: Each observation is the average share of parents who enroll in their first-best major choice as a function of their GPA. Each dot is a 0.1 GPA bin, except for the leftmost dot which is a 0.5 bin due to sparsity. The vertical lines denote the admissions GPA cutoff (normalized to 0). The number of observations in panel A is 37,390 (father-son sample), in panel B 49,057 (mother-son sample), in panel C 46,791 (mother-daughter sample), and in panel D 35,695 (father-daughter sample).

Figure A6. Comparison of the 2-slope versus 12-slope models.

Panel A: Siblings



Panel B: Intergenerational



Notes: The first column plots averages of the binned outcome variable for younger siblings and children against the running variable, allowing for separate slopes for each of the five first-best choices to the right of the cutoff and a common slope to the left of the cutoff. The second column shows similar plots, but allowing separate slopes for each of the seven second-best choices to the left of the cutoff and a common slope to the right of the cutoff. The trend lines are RD estimates using the underlying data, no covariates, and triangular weights. Note that these graphs are for illustrative purposes; we never mix the 2-slope and 12-slope models in estimation.

Table A1. Summary statistics for all applicants with a first-choice academic program 1977-1991.

	Oversubscribed programs	Non-impacted programs
Father age	45.77	46.02
Mother age	43.21	43.34
Father schooling	11.63	11.32
Mother schooling	11.25	10.83
Father earnings	5.77	5.75
Mother earnings	5.23	5.20
Foreing born parent	0.17	0.17
Foreign born	0.04	0.04
Female	0.52	0.51
Age in year of applying	16.00	15.99
GPA	3.86	3.94
Observations	263,878	221,397

Notes: Parent and child characteristics are measured in the year of application (the child's 16th year since birth). Years of schooling inferred from highest education level. Parents here refer to the parents of applicants during 1977-1991 (i.e., these are the grandparents of the children in our intergenerational sample). Parent earnings are measured between the ages of 37-39 and are converted to year 2016 US dollars using an exchange rate of 8.5 SEK to 1 USD.

Table A2. Comparison of major cutoffs across years within the same school region.

	Fraction of years with a higher cutoff		
Major combinations	1st major	2nd major	No difference
Engineering vs. Natural Science	.37	.25	.38
Engineering vs. Business	.28	.42	.30
Engineering vs. Social Science	.21	.53	.27
Engineering vs. Humanities	.31	.38	.31
Natural Science vs. Business	.24	.46	.30
Natural Science vs. Social Science	.18	.51	.31
Natural Science vs. Humanities	.24	.38	.39
Business vs. Social Science	.24	.48	.28
Business vs. Humanities	.37	.32	.31
Social Science vs. Humanities	.47	.21	.32

Notes: The table reports the average fraction of years with a higher cutoff for one major compared to another within the same school region. If both majors have a cutoff in a given year in the same school region, we compare the two to determine which is higher. If one major has a cutoff, but the other does not, we record the major with the cutoff as having a higher cutoff. "No difference" can either reflect that both majors have cutoffs which are equal or that neither major was oversubscribed.

Table A3. Correlational estimates by gender mix.

	Correlational estimates	IV-enrolled	Difference
Panel A: Siblings			
Older brother – younger brother	.182***	.063***	.119***
	(.004)	(.015)	(.014)
Older sister – younger brother	.056***	-.029**	.085***
	(.003)	(.014)	(.014)
Older sister – younger sister	.108***	.039**	.069***
	(.004)	(.016)	(.016)
Older brother – younger sister	.098***	.017	.081***
	(.003)	(.014)	(.014)
N	82,706	88,164	
Panel B: Intergenerational			
Father – son	.111***	.043***	.068***
	(.003)	(.012)	(.021)
Mother – son	.045***	.024**	.021**
	(.002)	(.011)	(.011)
Mother – daughter	.054***	.010	.044***
	(.003)	(.012)	(.012)
Father – daughter	.061***	.014	.047***
	(.003)	(.012)	(.011)
N	157,760	168,933	

Notes: Correlational estimates are based on the fraction of younger siblings/children who list a major as their first choice if it is the one their older sibling/parent enrolled in minus the fraction who choose it when their older sibling/parent did not enroll in it. This is done for each of the 5 majors and averaged across majors (with weights equal to the number of older siblings/parents choosing each of the majors). Bootstrap standard errors based on 1,000 replications. For IV-enrolled estimates, see notes to Table 5 and 6.

** p<0.10, ** p<0.05, *** p<0.01*

Table A4. Intergenerational estimates by birth order.

	Firstborn child			Not firstborn child		
	Reduced form	IV-enrolled	IV-completed	Reduced form	IV-enrolled	IV-completed
(1) Impact on sons						
Father – son	.040***	.062***	.078***	.017	.025	.033
	(.011)	(.017)	(.021)	(.011)	(.017)	(.022)
Mother – son	.015	.027*	.035*	.013	.019	.025
	(.010)	(.016)	(.019)	(.010)	(.015)	(.019)
(2) Impact on daughters						
Mother – daughter	.010	.020	.027	-.002	-.001	.002
	(.011)	(.017)	(.021)	(.011)	(.017)	(.022)
Father – daughter	.010	.021	.030	.004	.007	.011
	(.010)	(.016)	(.020)	(.010)	(.015)	(.020)
N	87,272	87,272	87,272	81,661	81,661	81,661

Notes: See notes to Table 6. Standard errors in parentheses, clustered at the family level.
** p<0.10, ** p<0.05, *** p<0.01*

Table A5. Alternative measures for whether a younger sibling or child copies their older sibling or parent.

Panel A: Siblings

	Baseline	Same major any rank	Same major accepted	Same major enrolled	Same major completed
(1) Reduced form					
Older brother – younger brother	.043***	.041***	.041***	.042***	.037***
	(.010)	(.010)	(.009)	(.009)	(.009)
Older sister – younger brother	-.026***	-.029***	-.024***	-.023***	-.018**
	(.009)	(.010)	(.008)	(.008)	(.008)
Older sister – younger sister	.025**	.025**	.030***	.029***	.027***
	(.011)	(.012)	(.010)	(.010)	(.010)
Older brother – younger sister	.009	.005	.007	.005	.004
	(.009)	(.010)	(.009)	(.009)	(.008)
(2) IV-enrolled					
Older brother – younger brother	.063***	.059***	.060***	.061***	.054***
	(.015)	(.015)	(.014)	(.014)	(.013)
Older sister – younger brother	-.029**	-.034**	-.028**	-.026**	-.020*
	(.014)	(.014)	(.012)	(.012)	(.011)
Older sister – younger sister	.039**	.038**	.045***	.044***	.040***
	(.016)	(.017)	(.015)	(.015)	(.014)
Older brother – younger sister	.017	.011	.014	.012	.009
	(.014)	(.015)	(.014)	(.014)	(.013)
N	88,164	88,164	88,164	88,164	88,164

Panel B: Intergenerational

	Baseline	Same major any rank	Same major accepted	Same major enrolled	Same major completed
(1) Reduced form					
Father – son	.029***	.029***	.021***	.019***	.023**
	(.008)	(.008)	(.007)	(.007)	(.010)
Mother – son	.015**	.015*	.010	.008	.004
	(.007)	(.008)	(.007)	(.006)	(.009)
Mother – daughter	.004	.001	.006	.005	.004
	(.008)	(.008)	(.008)	(.007)	(.010)
Father – daughter	.007	.009	.002	.002	-.007
	(.008)	(.008)	(.007)	(.007)	(.009)
(2) IV-enrolled					
Father – son	.043***	.045***	.031***	.028**	.033**
	(.012)	(.013)	(.011)	(.011)	(.015)
Mother – son	.024**	.024**	.017*	.014	.007
	(.011)	(.012)	(.010)	(.010)	(.013)
Mother – daughter	.010	.006	.012	.009	.007
	(.012)	(.012)	(.011)	(.011)	(.014)
Father – daughter	.014	.017	.006	.005	-.007
	(.012)	(.012)	(.011)	(.011)	(.014)
(3) IV-completed					
Father – son	.056***	.057***	.040***	.036**	.041**
	(.016)	(.017)	(.015)	(.014)	(.019)
Mother – son	.031**	.031**	.022*	.018	.010
	(.014)	(.015)	(.013)	(.012)	(.017)
Mother – daughter	.015	.011	.016	.013	.010
	(.015)	(.016)	(.014)	(.014)	(.018)
Father – daughter	.021	.024	.011	.008	-.006
	(.015)	(.015)	(.014)	(.014)	(.018)
N	168,933	168,933	168,933	168,933	93,412

Notes: See notes to Tables 5 and 6. The baseline outcome variable is whether a younger sibling's or child's first choice on their preference list matches the first-best major choice of their older sibling or parent. Columns 2-5 replace this with whether the younger sibling or child (i) includes on their choice list, (ii) is accepted to, (iii) enrolls in, or (iv) completes the same major as the first-best choice of their older sibling or parent. Standard errors in parentheses, clustered at the family level.
** p<0.10, ** p<0.05, *** p<0.01.*