NBER WORKING PAPER SERIES

FAMILY SPILLOVERS IN FIELD OF STUDY

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

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.

Family Spillovers in Field of Study
Gordon B. Dahl, Dan-Olof Rooth, and Anders Stenberg
NBER Working Paper No. 27618
July 2020
JEL No. I21,J24

ABSTRACT

This paper estimates peer effects both from older to younger siblings and from parents to children in academic fields of study. Our setting is secondary school in Sweden, where admissions to oversubscribed fields is determined based on a student's GPA. Using an RD design, we find strong spillovers in field choices that depend on the gender mix of siblings and whether the field is gender conforming. There are also large intergenerational effects from fathers and mothers to sons, except in female-dominated fields, but little effect for daughters. These spillovers have long-term consequences for occupational segregation and wage gaps by gender.

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

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

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

1 Introduction

A large literature documents the return to years of schooling (Card 1999). But only recently have researchers been able to show that the choice of *what* to study can play an equally important role in labor market outcomes, with traditionally male-dominated fields yielding higher earnings.[1] While little is known about these consequential field of study decisions, the influential Coleman report (1966) posited the family plays an outsized role in education, and argued for governments to adopt measures to counter any resulting inequality. Yet over 50 years later, causal evidence on the role of siblings and parents in field choice, and the gendered mechanisms through which such spillovers might operate, remains elusive.

Understanding the role of social interactions within the family has proven challenging for two reasons. The first problem is correlated unobservables.[2] Field choices could be driven by common factors such as family income or shared genes, rather than a peer effect. Second, it is difficult to access and link datasets which contain the field choices of siblings or of parents and their children. We overcome these challenges in the context of Sweden's secondary school system. As in many countries, students choose fields of study (i.e., "high school majors") which prepare them for college or provide vocational training (for an overview, see OECD 2019). These fields are often oversubscribed, with admission to fields rationed by GPA. We use high-quality register data of applicants from 1977-1991 and a regression discontinuity design to compare older siblings or parents just above versus just below admission cutoffs, and see (i) if their younger siblings or (ii) if their children (using data from 2011-2019) are affected in their field of study choice.

Starting with siblings, our empirical results show strong spillovers for those of the same gender. Younger brothers are 6 percentage points more likely to choose a field if their older brother enrolled in it, and a similar 3.5 percentage point increase is found for sister pairs. For mixed-gender siblings, there is a recoil effect for younger brothers, where they avoid their older sister's field. Older brothers have no detectable influence on younger sisters.

These patterns become even more intriguing when broken down by whether a field is

[1]See Altonji et al. 2012, Andrews et al. 2017, Dahl et al. 2020, Hastings et al. 2013, Kirkeboen et al. 2016.

[2]Other challenges to estimating peer effects include endogenous group membership and reflection (Manski 1993). Endogenous group membership is not an issue since individuals do not choose their siblings or children, and reflection is not an issue if older siblings and parents influence younger siblings and children, but not the other way around (note, however, that an experimental design does not require this assumption).

gender conforming (defined using the gender makeup of a field). Younger brothers follow their older brother into the male-dominated field of Engineering (8.4 pp effect) and also into the gender-balanced fields of Natural Science or Business (5.6 pp effect), but not into the female-dominated fields of Social Science or Humanities. In contrast, younger brothers turn away from fields which are female dominated if their older sister enrolls in them (-4.8 pp effect). Effects for sister pairs are present, but do not follow field gender norms.

Turning to parents and their children, we find strong intergenerational spillovers. On average, sons are more heavily influenced than daughters. If a father or mother completes a field, their son is 5.1 or 2.4 percentage points, respectively, more likely to choose the same field. But these effects mask substantial heterogeneity. Sons are strongly influenced by both parents to pursue Engineering, Natural Science, and Business, but not the female-dominated fields of Social Science and Humanities. For daughters, there is a hint that mothers who complete an Engineering degree provide a role model, but no effect from parents choosing other fields.

Taken together, our results indicate the family serves to reinforce gender stereotypical norms for males, but not for females. The resulting gendered choices have long-term labor market consequences. In recent work, we show these early field specializations affect adult earnings, with students who major in the female-dominated fields of Social Science or Humanities earning substantially less (Dahl et al. 2020). The earnings losses are as large as the return to two additional years of schooling. Most of this difference in adult earnings can be attributed to differences in subsequent college major and occupation.

The sibling and intergenerational results complement each other well. Because we use the same time period for both parents and older siblings, we are uniquely able to compare spillover magnitudes. We find that an individual's field choice has a larger influence on their siblings than their children. The sibling spillovers occur relatively close in time, when curricula and information about a field 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 boys being influenced in their field choices by family members, but only if the field is not overly populated with females, consistent with models of gender identity (Akerlof and Kranton 2000). Girls are less influenced by family members, although some role model effects appear for mothers and older sisters who choose fields which are not dominated by females.

Our paper is related to three strands of research: family spillovers in years of schooling, choice of college, and course content.[3] To study the causal link between parent's and children's years of schooling, researchers have used school reforms, twins, and adoptions, often reaching different conclusions on whether there is an intergenerational link (for reviews, see Black and Devereux 2010, Björklund and Salvanes 2011, Holmlund et al. 2011). In contrast, there are not similar causal studies for siblings.

To study the college choices of siblings, a recent set of papers have used RD designs based on GPA cutoffs. Altmejd et al. (2020) 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 no extensive margin response, but compelling evidence that younger siblings follow their older siblings to the same college. Aguirre and Matta (2018) and Dustan (2018) find evidence for similar types of sibling spillovers in Chile and Mexico. None of these recent papers finds direct evidence for spillovers in field of study; this is largely due to data limitations given the way college admissions works in these countries.[4] There are no similar studies for causal parent-child spillovers in college choice.

To study 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.[5] Their paper is notable in providing some of the first

[3]There is also a literature on how individuals more generally choose which field to study (Wiswall and Zafar 2015, Zafar 2013).

[4]In the U.S., individuals apply to each college individually, without taking into account field. In contrast, in Chile, Sweden, and Croatia, students apply to a nationally centralized administration that admits students based on their ranking list and their GPA. These preference rankings indicate a combined college plus field of study. For example, a student's ranking might be 1. Stockholm University+Engineering, 2. Gothenburg University+Engineering, and 3. Stockholm University+Business. Altmejd et al. (2020) call the combined college plus field of study a "major choice", but their labelling should not be confused with field of study. As might be expected, students often prioritize field of study over college institution, and so there is more identifying variation for college choice (holding field fixed). One advantange of our setting is that students are only choosing fields of study, and not which secondary school to attend during our time period.

[5]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. See Table 4 in Joensen and Nielsen (2018).

causal evidence of sibling spillovers in educational content.

Our paper makes several contributions relative to this literature.[6] While years of schooling, college choice, and course content are all important choice margins, field of study captures a different type of family spillover and has important labor market consequences. Our first contribution is that we find novel evidence for sibling peer effects in field of study, in a setting where there is no confound of which school to choose. Using a convincing RD design and 15 years of data, we show these sibling influences depend on sibling gender composition and whether a field is gender conforming in ways which have implications for future college major, occupation, and earnings. Second, to our knowledge, we are the first to causally estimate parent-to-child spillovers in field of study using an intergenerational panel dataset. We document that both father's and mother's field choices matter for sons roughly 30 years later, unless the field was dominated by females, but that there are few spillovers for daughters.

Returning to the Coleman report, our results substantiate that families play a key role in education outcomes through the channel of field choice. These family spillovers will amplify the effects of education policies. One reform is to expand slots in the higher-paying fields of Engineering, Natural Science, and Business. But a gender-blind expansion will also reinforce traditional gender norms for boys. More targeted policies, such as nudging girls into STEM or Business fields through outreach campaigns, or making it easier for girls to gain admission to these higher-paying fields, would help decrease gender occupation segregation and earnings gaps in the long run.[7]

2 Setting and Data

2.1 Admission to Fields of Study in Sweden

After nine years of compulsory schooling, individuals in Sweden can apply to a field of study in secondary school. In this paper, we focus on the five academic fields which are preparatory for university studies: Engineering, Natural Science, Business, Social Science,

[6]More broadly, there is a literature which uses observational designs or studies different outcomes. For a sampling of recent papers, see Anelli and Peri 2015, Black et al. 2017, Breining et al. 2015, Cools and Patacchini 2017, Nicoletti and Rabe 2019, Quereshi 2018.

[7]For example, Breda et al. (2020) and Porter and Serra (2020) find that exposure to female role models increases enrollment in STEM and Economics majors, respectively (see also Buckles 2019). Teacher gender has also been shown to matter (Bettinger and Long 2005, Carrell et al. 2010). A different approach is to give females an advantage when applying to male-dominated fields, a policy Sweden has pursued (see Section 2.1).

and Humanities. During the periods we study, roughly half of students choose one of these fields. Since non-academic fields are usually not oversubscribed, we include them only as an outside option.

Students compete for slots based on their application GPA, which is the average grade across 10-12 school subjects as of ninth grade. GPA has an approximate mean of 3 and standard deviation of 1 in the entire population. Applicants have an extra 0.2 added to their GPA if they apply to a field which accepted 30% or less of their gender nationally in the prior year. Students rank their preferences for up to 6 fields, and the central administration office then allocates students. Admission decisions are made sequentially, with the highest-GPA applicant being admitted to their first-choice field, the second-highest GPA applicant being admitted to their highest-ranked field among the set of fields 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).[8]

After admission decisions are sent out in July, there can be reshufflings across different fields as student 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 can be used in an RD design to instrument for either field enrollment or completion. Appendix A describes how these cutoffs are determined in our data.

2.2 Data

We have information on the field ranking lists submitted by all students going back to 1977. This allows us to not only observe which field an applicant is admitted to, but also to account for next-best alternatives. This is key for being able to identify an interpretable causal effect (see Kirkeboen et al. 2016). If an individual is admitted to either their first or second ranked choice, which happens 95% of the time, then we define these as the individual's preferred and next-best alternative fields, respectively. For individuals who are admitted to a third or lower ranked choice, their preferred choice is the field ranked immediately above their accepted

[8]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 given for the first and second choices on a ranking list, so individuals may not have revealed their true preferences. As we show, excluding these years does not materially affect our estimates.

choice, and the next-best alternative is their accepted choice. For ease of exposition, in what follows we will refer to the preferred field 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 fields of study did not experience major changes and students were only choosing fields and not which school to attend. In 1992, Business, Social Science, and Humanities were merged into one field, only to re-emerge as separate fields again in 2011. The introduction of private schools and school choice, as well as other reforms, also substantially reduced the number of oversubscribed fields after 1992. This means 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 (see Appendix A). 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.

Figure A1 shows the distribution of field 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 field choices, with over one-fourth of applicants choosing each of these fields. 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 fields 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 fields induce different types of family spillover patterns.

Our estimation samples for older siblings and parents are limited to those who have a first-best academic choice where demand exceeds supply. During 1977-1991, 55% of individuals apply to a first-best choice which has competitive admissions. 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 fields were combined. Finally, we cap family size at 5 siblings or 5 children, which drops 1.6% and 0.5% of the data, respectively. This leaves us with a sample of 88,299 sibling pairs and 169,203 parent-child pairs.

3 Model

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

Our goal is to estimate spillover effects in field choices. 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 fields of study, but also an accounting of second-best choices. A randomly assigned cutoff for each parent's field in a fuzzy RD design will eliminate selection bias, but without further modifications and/or assumptions, it will not estimate the intergenerational effect for any particular group of parents who complete one field over another.[9]

When next-best alternatives are available, RD can estimate LATEs for every combination of preferred and next-best fields. One additional necessary assumption is what Kirkeboen et al. label the "irrelevance condition." In our context, this condition means that if crossing the GPA threshold for admission to Engineering does not cause a parent to complete Engineering, then it does not cause the parent to complete another field like Humanities either.

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 most general reduced form RD equation can written as:

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

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 field as a parent who had preferred field j and next-best alternative k. The running variable x is the parent's GPA, c_j is the cutoff GPA for admission to field j, g^l_{jk} and g^r_{jk} are unknown functions

[9] For example, estimates would not be informative about whether spillovers for parents choosing Engineering instead of Social Science are larger or smaller than for those choosing Business instead of Social Science.

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 θ_{jk} coefficients capture whether a child is more likely to copy the choice of a parent who is admitted to field j instead of their next-best alternative k. We combine data across years and school regions by normalizing each cutoff to be 0, and adjust the GPA running variable accordingly. We estimate fuzzy RDs using either a parent's enrollment in or completion of a field in first stage regressions.

Equation (1) is extremely flexible in how preferred and next best alternatives enter the regression. In practice, we need to make several simplifying parametric assumptions. First, we assume that the peer effect coefficient θ_{jk} does not depend on k, so that a parent's next best choice does not directly influence the child's choice. Our baseline model also parameterizes $\alpha_{jk} = \delta_j + \tau_k$, so that instead of 30 different intercept terms, we allow for 5 different intercepts based on first choices and 7 based on second choices.[10]

Next, we limit the number of functions g^r_{jk} and g^l_{jk} which need to be estimated. 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). We find that the 2-slope model can be too restrictive, but that the 12-slope and 60-slope models reach similar conclusions.

It is useful to compare the parameterizations we make to related studies. In terms of our notation, Kirkeboen et al. (2016) imposes a similar restriction as we do that $\theta_{jk} = \theta_j$, parameterizes $\alpha_{jk} = \delta_j + \tau_k$, but controls for the running variable with a single common slope. Altmejd et al. (2020) instead assumes that $\theta_{jk} = \theta$, parameterizes $\alpha_{jk} = \delta_j + \tau_k$, and uses a two-slope model. So while we make some parametric assumptions due to sample size considerations, they are less restrictive than those made by other researchers using similar designs and settings.

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

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 manipulation in our setting. One reason is that the required GPA is not known in advance, and varies from year to year as a function of the number of applicants. Thresholds differ over 80% of the time for fields with a cutoff in successive years, as shown in Figure A2. One way to test for manipulation is to check whether pre-determined characteristics are balanced around the admission cutoff. We do this in Figure A3 for all applicants to a competitive field of study during 1977-1991. There are no noticeable jumps at the cutoff, a finding which is confirmed with formal tests which are not shown.[11]

To identify the causal spillover effects of enrollment or completion, we additionally need exclusion, monotonicity, and irrelevance conditions which are not required for the reduced form. The exclusion restriction requires that crossing the admissions threshold for a field only affects outcomes through enrollment (or completion). For enrollment, being admitted but not enrolling provides little information or experience with a field. For completion, it is possible that parents take some specialized courses because they are admitted to a field, but then do not complete the field. Given that most switching occurs early on during the fall of the first year, however, there is limited scope for this channel. The monotonicity assumption requires that crossing an admissions threshold does not make an individual less likely to complete that field, which seems plausible. Finally, we require the irrelevance condition discussed in Section 3.1. This condition also seems reasonable in our setting.

4 Results

4.1 First Stages

The first stage for older siblings is how admissions affects field enrollment, while for parents we can use either field enrollment or field completion. We cannot use field completion for siblings because the older sibling often has not completed their secondary school field before the younger sibling makes their field choices.

Figure 1 illustrates the first stages, separated by gender mix, for a two slope model. Consider the top-left graph, which plots the probability an older brother enrolls in his

[11] As discussed by Dahl et al. (2020), a standard McCrary (2008) test or the newer density test proposed by Cattaneo, Jansson, and Ma (2018) is not applicable since the cutoff is based on an order statistic.

first-best choice in normalized GPA bins, for the sample with a younger brother. Everyone to the right of the vertical line is (initially) admitted to their first-best field, while everyone to the left is not (initially) admitted. Some people switch to other fields, usually before the school year begins. This reshuffling opens up slots for other students and explains why some individuals to the left of the admissions cutoff are able to enroll in their first-best choice.[12] Switching fields is not necessary random which is why we need to instrument for enrollment with admission. There is a roughly 60% jump in the probability of enrollment at the cutoff; similarly-sized jumps occur for the other sibling gender combinations (top 4 graphs).

The graphs for parental enrollment look similar to those for siblings, and are therefore relegated to the appendix (Figure A4). Instead, the bottom four graphs of Figure 1 plot the probability a parent *completes* their first-best choice. A nontrivial fraction of those admitted to their first-best field do not complete it. This is mostly due to individuals switching to other fields, as few drop out entirely. The four graphs are similar to each other, with a roughly 45% jump in the probability a parent completes their first-best choice at the cutoff.

The bottom half of Table 1 reports first stage regression estimates. They are all highly significant, indicating there will not be a weak instrument problem.

4.2 Sibling Spillovers in Field of Study

Estimates by Gender Mix. Estimates of sibling spillovers are reported in the upper left panel of Table 1. The dependent variable is a dummy for whether a younger sibling's field choice, defined as the first choice on their list of preferences, is the same as their older sibling's first-best choice. We explore alternative outcome measures in Section 4.4 (see Table A2). To gain precision, we stack the data for different gender-mix combinations into a single regression, and include gender mix times field specific dummies for each first best-choice (20 dummies), dummies for each second best choice (7 dummies), common year and time fixed effects, and use the linear 12-slope model with triangular weights as discussed in Section 3. Standard errors are clustered at the family level. Corresponding graphs for the reduced form effects can be found in Figure 2.

There are strong spillover effects, but which vary by gender mix. Start with brother pairs.

[12]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 share to the right of the cutoff can explain a large increase in the enrollment share to the left of the cutoff.

The reduced form shows that if an older brother is admitted to his first-best choice, there is a 4.2 percentage point higher probability his younger brother will choose the same field. The corresponding IV estimate implies a younger brother is 6 percentage points more likely to choose the same field as the one his older brother enrolled in. Turning to sister pairs, there is also a strong same-sex sibling effect, although not as large. An older sister's enrollment increases the chances her younger sister will choose the same field by 3.5 percentage points. To put the magnitude of the IV estimates in perspective, the 6 and 3.5 percentage point effects, respectively, translate into 23% and 16% increases relative to their means near the cutoff (see Figure 2). Older sisters' peer influence is stronger when siblings are closer in age, while older brothers are more influential when farther apart (see Table A1).

We next examine the spillover effects for opposite-gender sibling pairs. Which field an older brother enrolls in has no discernable impact on his younger sister's choice. But the same is not true for older sisters and their younger brothers. In these families, younger brothers recoil from the fields of their older sisters. A younger brother is 3.6 percentage points *less* likely choose a field if his older sister enrolls in it.

While this heteregeneity 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, we estimate an IV coefficient of 2.1 percentage points which is marginally significant.

Estimates by Gender-Conforming Fields. The results become even more intriguing when we estimate separate peer effects by the share of females in a field. We continue to estimate the effect of a younger sibling choosing exactly the same field as their older sibling. But instead of imposing a common treatment effect, we estimate separate effects by field type. We define a male-dominated field as Engineering (less than 20% female), female-dominated fields as either Social Science or Humanities (both over two-thirds female), and gender-balanced fields as Natural Science or Business (see Figure A1).

The results using this split are found in panel A of Table 2. Begin with brother pairs. Brothers copy brothers for the heavily male-dominated field of Engineering (8.4 pp effect) and also into the gender-neutral fields of Natural Science and Business (5.6 pp effect). But younger brothers do not appear to follow older brothers into the gender-nonconforming fields of Social

Science or Humanities. Sister pairs, in contrast, do not exhibit such gender-conforming patterns. Sisters copy sisters into gender-neutral fields, but not the female-dominated fields. There is a large effect for Engineering, but it occurs so rarely as to be statistically insignificant.

The patterns are equally interesting for opposite-gender siblings. Younger brothers are less willing to choose a female-dominated field if their older sister enrolled in it. There is a 4.8 percentage point drop in the likelihood a younger brother chooses Social Science or Humanities if this would copy his sister. In contrast, older brothers have no noticeable effect on their younger sisters, regardless of the share of females in a field. 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.

4.3 Intergenerational Spillovers in Field of Study

Estimates by Gender Mix. The upper right panel of Table 1 reports estimates for parents and their children, broken down by gender mix (see also Figure 2). We use the same regression specification as we did for siblings.

We find sizable spillovers from fathers to sons and a hint of spillovers from mothers to sons. There is a 2.8 percentage point increase in a son's probably of choosing a field if his father was admitted to it, a 4 percentage point increase if his father enrolled in it, and a 5.1 percentage point increase if his father completed it. The IV estimates based on enrollment and completion, respectively, translate into 16% and 20% increases relative to the mean near the cutoff (see Figure 2). For mothers and sons, the equivalent effect sizes are somewhat smaller, and on the border of statistical significance. As Table A1 shows, there is some evidence that a father's effect on his son is somewhat larger for first-borns. In contrast, we find little effect of parents on daughters. The estimates for father-daughter and mother-daughter pairs are both relatively small and not statistically significant.

Table 1 also reports a combined estimate which ignores the gender composition of a parent-child pair. Estimating a common effect leads to a misleading conclusion, since sons are strongly affected, but daughters less so.

Estimates by Gender-Conforming Fields. In Table 2 we further break down the estimates by whether the child chooses a gender-conforming field. For sons, we find strong peer effects from both fathers and mothers, but only if the field is not female dominated. Sons are 5 and

6.5 percentage points, respectively, more likely to choose the field their *father* completed if it was male dominated or gender neutral. Likewise, sons are 18.9 and 3.8 percentage points, respectively, more likely to copy their *mother* for male-dominated and gender-neutral fields. In contrast, for female-dominated fields there is no statistical evidence that sons copy parents.

Turning to daughters, there is little evidence for spillovers from either fathers or mothers. The lone exception is a hint of mothers being a role model for Engineering, an estimate which is significant at the 10% level.

4.4 Specification Checks

Table 3 presents a variety of robustness checks. Since the reduced form and IV estimates show similar patterns, we focus on the former.

We start by probing the parameterization of the regression model. 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%.[13] Column 3 reveals the results are robust to cutting the window size in half. We next show both the more parsimonious 2-slope model and the more flexible 60-slope model. The results are similar to baseline, except that for the 2-slope model the older-younger sister estimate is no longer statistically significant. In Figure A5, we illustrate why the 2-slope model is too restrictive. For example, the slope for Natural Science is steeper to the right of the cutoff compared to other fields for the parent-child graph in panel A. In columns 6 and 7 we omit 1982-84 (see footnote 8) and add in predetermined demographic variables. The results remain similar to baseline.

In the final column of Table 3, we estimate a placebo specification. Specifically, we test whether a younger sibling's acceptance into a field affects their older sibling's choices. Since the older sibling makes their choices before their younger sibling, there should not be an effect. Indeed, the placebo estimates are close to zero and not statistically significant. A similar exercise cannot be performed for parents and children (see Appendix A).

Finally, in Table A2, we explore alternative measures for whether a younger sibling or child copies their older sibling or parent. Our baseline definition used the first-choice field on a younger sibling's or child's ranking list. If we instead use whether they listed the field at

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

all, 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 field as their older sibling, and become slightly smaller if we use field completion. For the intergenerational analysis, the magnitudes drop by roughly 30-40% if we use acceptance, enrollment, or completion of the same field. These patterns hold for both the reduced form and IV specifications.

5 Conclusion

Our findings provide novel evidence that families play an important role in the field choices individuals make at an early age, in ways that have long-term labor market consequences. 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 field is gender conforming. Our results confirm theories and predictions that sibling and parent-child dynamics are differentially influenced by gender mix (for summaries, see Bush and Peterson 2013, McHale et al. 2013, Murry et al. 2013). Interestingly, we find that an individual's field choice has a larger influence on their sibling than their child, despite the literature's larger focus on parental influences (McHale et al. 2012). The patterns we observe also emphasize that families play a key role in the development of gender norms, a pathway highlighted by Bertrand (2011) and Brenøe (2018).

References

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

Akerlof, G. A. and R. E. Kranton (2000): "Economics and identity," *The Quarterly Journal of Economics*, 115, 715–753.

Altmejd, A., A. Barrios-Fernández, M. Drlje, J. Goodman, M. Hurwitz, D. Kovac, C. Mulhern, C. Neilson, and J. Smith (2020): "O brother, where start thou? Sibling spillovers on college and major choice in four countries," CEP Discussion Paper No. 1691.

Altonji, J. G., E. Blom, and C. Meghir (2012): "Heterogeneity in human capital investments: High school curriculum, college major, and careers," *Annual Review of Economics*, 4, 185–223.

Andrews, R. J., S. A. Imberman, and M. F. Lovenheim (2017): "Risky business? The effect of majoring in business on earnings and educational attainment," NBER Working Paper No. 23575.

Anelli, M. and G. Peri (2015): "Gender of siblings and choice of college major," *CESifo Economic Studies*, 61, 53–71.

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.

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 (2017): "Sibling spillovers," NBER Working Paper No. 23062.

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, C. Van Effenterre, et al. (2020): "Do female role models reduce the gender gap in science? Evidence from French high schools," IZA Discussion Paper No. 13163.

Breining, S., N. M. Daysal, M. Simonsen, and M. Trandafir (2015): "Spillover effects of early-life medical interventions," IZA Discussion Paper No. 9086.

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.

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

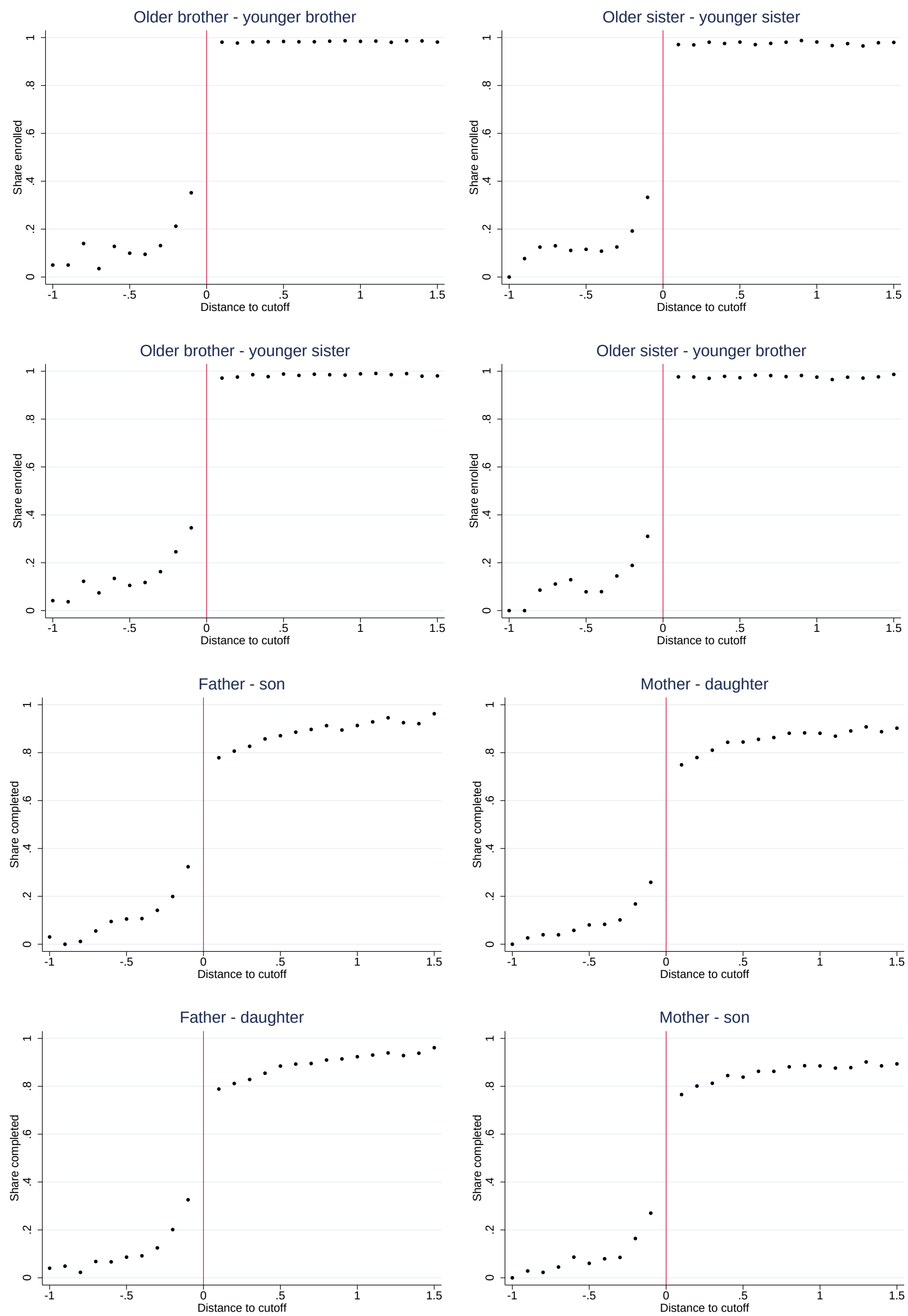
Card, D. (1999): "The causal effect of education on earnings," in *Handbook of labor economics*, Elsevier, vol. 3, 1801–1863.

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

Cattaneo, M. D., M. Jansson, and X. Ma (2018): "Manipulation testing based on

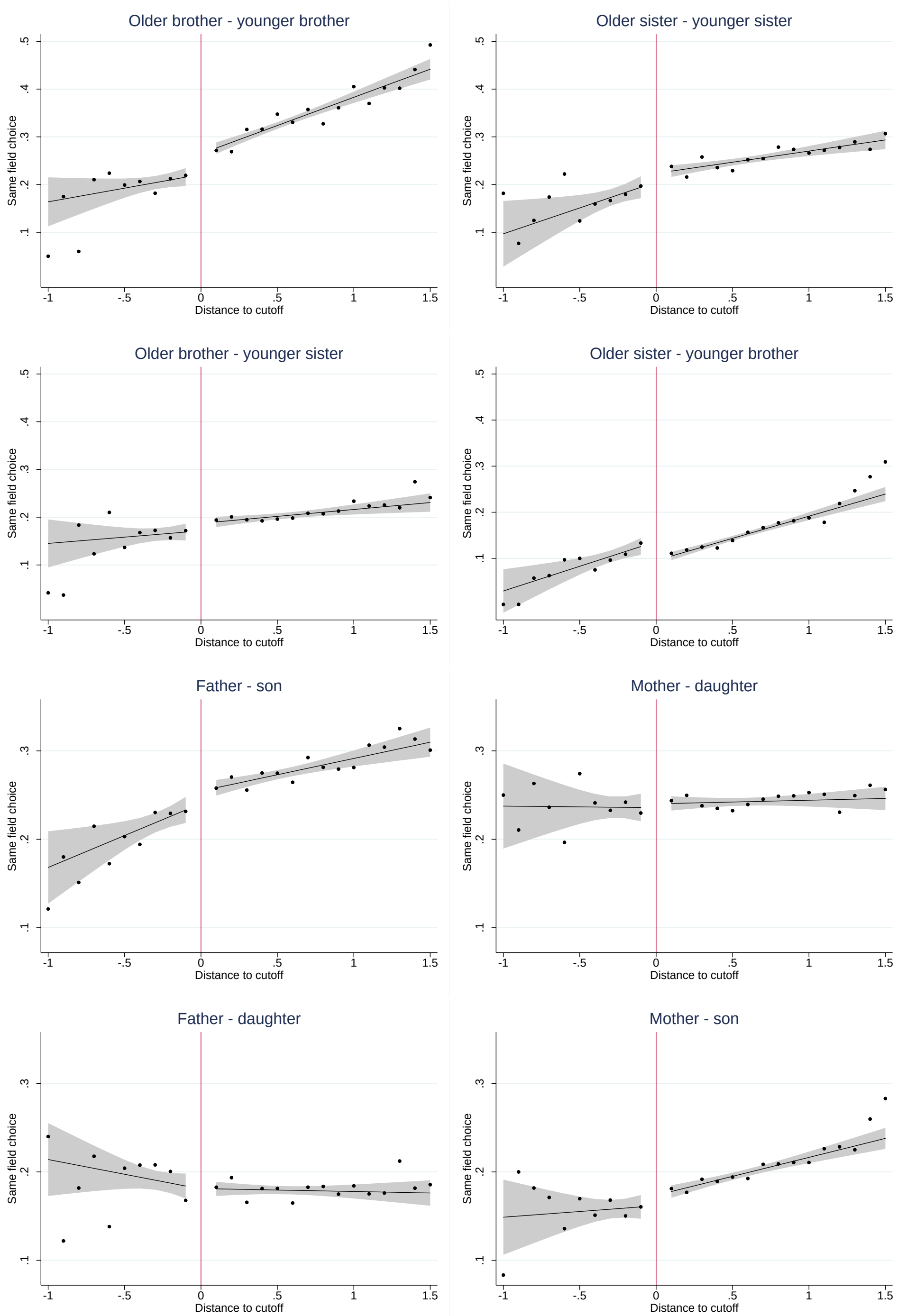
density discontinuity," *The Stata Journal*, 18, 234–261.
Coleman, J. S., E. Campbell, C. Hobson, J. McPartland, A. Mood, F. Weinfeld, and R. York (1966): "The Coleman report," *Equality of Educational Opportunity*.
Cools, A. and E. Patacchini (2017): "Sibling gender composition and women's wages," IZA Discussion Paper No. 11001.
Dahl, G. B., D.-O. Rooth, and A. Stenberg (2020): "Long-run returns to field of study in secondary school," NBER Working Paper No. 27524.
Dustan, A. (2018): "Family networks and school choice," *Journal of Development Economics*, 134, 372–391.
Hastings, J. S., C. A. Neilson, and S. D. Zimmerman (2013): "Are some degrees worth more than others? Evidence from college admission cutoffs in Chile," NBER Working Paper No. 19241.
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.
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," *The Quarterly Journal of Economics*, 131, 1057–1111.
Manski, C. F. (1993): "Identification of endogenous social effects: The reflection problem," *The 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.
Murry, V. M., L. S. Mayberry, and C. Berkel (2013): "Gender and family relations," in *Handbook of Marriage and the Family*, Springer, 401–422.
Nicoletti, C. and B. Rabe (2019): "Sibling spillover effects in school achievement," *Journal of Applied Econometrics*, 34, 482–501.
OECD (2019): *Education at a Glance 2019: OECD Indicators*, Paris: OECD Publishing.
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.
Svensson, L.-G. (1999): "Strategy-proof allocation of indivisible goods," *Social Choice and Welfare*, 16, 557–567.
Wiswall, M. and B. Zafar (2015): "Determinants of college major choice: Identification using an information experiment," *The Review of Economic Studies*, 82, 791–824.
Zafar, B. (2013): "College major choice and the gender gap," *Journal of Human Resources*, 48, 545–595.

Figure 1. First stage: Share enrolling in their first-choice field (older siblings) and share completing their first-choice field (parents), as a function of normalized GPA.



Notes: Each observation is the average share of older siblings who enroll in their first-best field choice (top 4 graphs) or the average share of parents who complete their first-best field choice (bottom 4 graphs), as a function of their GPA. The vertical lines denote the admissions GPA cutoff (normalized to 0). The number of observations is 22,875 (older brother - younger brother), 21,057 (older sister - younger sister), 21,411 (older brother - younger sister), 22,956 (older sister - younger brother), 37,447 (father-son), 46,869 (mother-daughter), 35,754 (father-daughter), and 49,133 (mother-son).

Figure 2. Reduced form: Probability of choosing the same field as an older sibling or parent.



Notes: Each observation is the average share of younger siblings or children whose first choice on their preference list matches the first-best field choice of their older sibling or parent, as a function of their older sibling's or parent's GPA. The vertical line denotes the admissions GPA cutoff for older siblings or 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. Shaded area indicates the 95% pointwise confidence intervals.

Table 1. Reduced form and IV estimates of the probability of choosing the same field as an older sibling or parent.

Panel A: Siblings			Panel B: Intergenerational			
	RF	IV (enrolled)		RF	IV (enrolled)	IV (completed)
(1) All			**(1) All**			
Older sibling – younger sibling	.013*	.020*	Parent – child	.011**	.018**	.024**
	(.007)	(.011)		(.006)	(.009)	(.012)
(2) By gender mix			**(2) By gender mix**			
Older brother – younger brother	.042***	.060***	Father – son	.028***	.040***	.051***
	(.010)	(.015)		(.008)	(.012)	(.015)
Older sister – younger sister	.023**	.034**	Mother – daughter	.003	.007	.011
	(.011)	(.016)		(.008)	(.012)	(.015)
Older brother – younger sister	.007	.014	Father – daughter	.002	.005	.010
	(.009)	(.014)		(.008)	(.012)	(.014)
Older sister – younger brother	-.030***	-.036***	Mother – son	.012	.019*	.024*
	(.010)	(.014)		(.007)	(.011)	(.014)
		First stage			First stage	First stage
(1) All			**(1) All**			
Older sibling – younger sibling		.632***	Parent – child		.621***	.465***
		(.008)			(.007)	(.007)
(2) By gender mix			**(2) By gender mix**			
Older brother – younger brother		.633***	Father – son		.620***	.449***
		(.011)			(.009)	(.009)
Older sister – younger sister		.636***	Mother – daughter		.629***	.476***
		(.012)			(.009)	(.009)
Older brother – younger sister		.617***	Father – daughter		.609***	.457***
		(.011)			(.009)	(.009)
Older sister – younger brother		.648***	Mother – son		.626***	.480***
		(.012)			(.009)	(.009)
N	88,299	88,299		169,203	169,203	169,203

*Notes: The outcome variable is whether a younger sibling's or child's first choice on their preference list matches the first-best field 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 field, as a function of whether their GPA exceeded the admissions cutoff. IV (completed) uses whether the parent completed their first-best field. 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 field interacted with sibling/parent-child gender mix; and dummies for next-best alternative field. Standard errors in parentheses, clustered at the family level. * p<0.10, ** p<0.05, *** p<0.01*

Table 2. Estimates as a function of the gender makeup of fields.

Panel A: Siblings			Panel B: Intergenerational			
	RF	IV (enrolled)		RF	IV (enrolled)	IV (completed)
Older brother – younger brother:			**Father – son:**			
Male dominated field (E)	.058***	.084***	Male dominated field (E)	.019	.032*	.049*
	(.016)	(.023)		(.012)	(.019)	(.026)
Gender neutral field (N+B)	.040***	.056***	Gender neutral field (N+B)	.040***	.054***	.064***
	(.015)	(.020)		(.011)	(.015)	(.018)
Female dominated field (S+H)	.011	.017	Female dominated field (S+H)	.016	.024	.033
	(.016)	(.021)		(.015)	(.020)	(.027)
Older sister – younger sister:			**Mother – daughter:**			
Male dominated field (E)	.048	.085	Male dominated field (E)	.039	.085	.100*
	(.049)	(.082)		(.027)	(.057)	(.056)
Gender neutral field (N+B)	.037**	.052**	Gender neutral field (N+B)	.003	.007	.011
	(.017)	(.023)		(.011)	(.015)	(.018)
Female dominated field (S+H)	.009	.015	Female dominated field (S+H)	.000	.004	.007
	(.014)	(.020)		(.011)	(.016)	(.020)
Older brother – younger sister:			**Father – daughter:**			
Male dominated field (E)	-.014	-.017	Male dominated field (E)	.003	.008	.022
	(.011)	(.017)		(.008)	(.013)	(.017)
Gender neutral field (N+B)	.024	.034	Gender neutral field (N+B)	-.007	-.006	-.003
	(.015)	(.021)		(.012)	(.016)	(.019)
Female dominated field (S+H)	.020	.030	Female dominated field (S+H)	.018	.026	.034
	(.021)	(.027)		(.017)	(.023)	(.030)
Older sister – younger brother:			**Mother – son:**			
Male dominated field (E)	.028	.043	Male dominated field (E)	.083*	.177*	.188**
	(.069)	(.099)		(.046)	(.103)	(.095)
Gender neutral field (N+B)	-.025*	-.030	Gender neutral field (N+B)	.022**	.031**	.037**
	(.014)	(.020)		(.010)	(.014)	(.017)
Female dominated field (S+H)	-.040***	-.048***	Female dominated field (S+H)	-.000	.002	.005
	(.011)	(.015)		(.010)	(.014)	(.018)
N	88,299	88,299		169,203	169,203	169,203

*Notes: See notes to Table 1. The outcome variable remains whether a younger sibling's or child's first choice on their preference list exactly matches the first-best field choice of their older sibling or parent. The regressions differ by allowing for heterogeneous effects based on the gender makeup of fields. Fields 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 3. Reduced form estimates using alternative specifications.

	Baseline	Quadratic	Smaller Bandwidth	2-slope model	60-slope model	Excluding 1982-84	Including demographics	Placebo
Panel A: Siblings								
Older brother – younger brother	.042***	.045***	.041***	.043***	.043***	.042***	.047***	-.001
	(.010)	(.013)	(.012)	(.010)	(.010)	(.012)	(.013)	(.010)
Older sister – younger sister	.023**	.026*	.025**	.017	.024**	.028**	.026*	.007
	(.011)	(.014)	(.013)	(.011)	(.012)	(.014)	(.014)	(.011)
Older brother – younger sister	.007	.011	.009	.009	.010	.011	.012	-.000
	(.009)	(.012)	(.011)	(.009)	(.010)	(.011)	(.012)	(.009)
Older sister – younger brother	-.030***	-.027**	-.025**	-.037***	-.029***	-.029**	-.026**	-.013
	(.010)	(.013)	(.011)	(.009)	(.010)	(.012)	(.013)	(.009)
N	88,299	88,299	62,302	88,299	88,299	62,477	88,299	91,845
Panel B: Intergenerational								
Father – son	.028***	.036***	.028***	.029***	.026***	.038***	.028***	
	(.008)	(.010)	(.009)	(.008)	(.008)	(.009)	(.008)	
Mother – daughter	.003	.011	.009	-.004	.003	.004	.003	
	(.008)	(.010)	(.009)	(.008)	(.008)	(.009)	(.008)	
Father – daughter	.002	.010	.006	.003	.000	.005	.002	
	(.008)	(.010)	(.009)	(.008)	(.008)	(.009)	(.008)	
Mother – son	.012	.020**	.019**	.006	.012	.019**	.012	
	(.007)	(.010)	(.008)	(.007)	(.008)	(.009)	(.007)	
N	169,203	169,203	122,493	169,203	169,203	122,679	169,203	

*Notes: See notes to Table 1 for baseline specification. Column 2 includes quadratic functions of the running variable on both sides of the cutoff and column 3 reduces the bandwidth in half. Columns 4 and 5 use the 2-slope and 60-slope models. Column 6 excludes the years 1982-84, when bonus GPA points were added for the first and second choices on an individual's ranking list. Column 7 includes as additional controls the demographic characteristics appearing in Table A3 plus age at application. Column 8 is a placebo specification where we estimate whether a younger sibling's admission into a field affects their older sibling's field choice. Standard errors in parentheses, clustered at the family level. * p<0.10, ** p<0.05, *** p<0.01.*

Appendix A (for Online Publication)

Determining GPA cutoffs[1]

While we observe the choice rankings for each individual and the admission decision, the GPA cutoff is not recorded in the dataset. Instead, we must infer the GPA cutoff from the data ourselves. Fortunately, in most cases this is simple and transparent, as the rules appear to have been followed.

Each combination of year, region, and field has the potential to be a competition for slots. We refer to these as "cells." Our empirical design only applies to 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 field and everybody below the cutoff is not.[2]

One wrinkle 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.[3] 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

[1]This section is based on related work by Dahl et al. (2020).

[2]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 field 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 0.34 percent of the data. We also require there be at least 25 applicants and 3 observations to the left of the cutoff.

[3]GPA is only recorded to the first decimal.

define the GPA cutoff as 3.35.

To allow us to pool the data across regions and years, we normalize the cutoff GPA to 0. The modal cutoff GPA of 3.2 corresponds to roughly the 15th percentile of GPAs in our baseline sample of applicants to competitive academic fields. To put this in perspective, the modal cutoff GPA also corresponds roughly to the median GPA of all ninth graders (including those applicants not in our sample which have a preferred non-academic choice and also those who don't apply to secondary school at all).[4] While the cutoffs vary substantially, they generally are only binding for applicants with GPAs in the bottom half of our estimation sample. Both of these comparisons will be important to keep in mind when interpreting the estimates, which will capture local average treatment effects for applicants around the cutoff.

One thing to note about the admission cutoffs is that fields which attract the highest GPA individuals do not necessarily have the highest cutoffs, or even a cutoff at all. This is because the cutoff is determined by supply versus demand for a field. For example, average GPAs are highest in Natural Science, but in many cases all students are admitted because there are fewer applicants than slots. The fields most likely to be oversubscribed in our data are Engineering, Business, and Social Science.

For our key sample period of 1977-1991, we find that over half of students choose a field which is oversubscribed (and for which we identify a cutoff). During this period, students were only choosing field of study and were assigned which school to attend in their local area. In the 1990s, there were several reforms. One change was a restructuring of the fields, which saw Business, Social Science, and Humanities merge, only to re-enter as separate fields again in 2011. Another change was allowing the entry of publicly-funded private schools and introducing school choice, which meant that students chose a combination of institution plus field of study. For example, in this later period, a student's ranking list might be 1. School A+Engineering, 2. School B+Engineering, 3. School A+Business.

These reforms mean that for the sample period 2011-2019, we can accurately measure which field is part of a child's first choice (since we have access to their ranking list), and whether it matches their parent's field. But there are few field-specific cutoffs for this period, which means that we cannot feasibly study sibling spillovers. The first reason is that the

[4]The median unadjusted GPA is 3.2 for the sample of all ninth graders in the years 1988-1991. This is the first set of years when registers of GPA for the full population, including those not applying to secondary school, was first collected.

number of oversubscribed fields was dramatically reduced, because private schools tended to establish in regions where there was excess demand (by 2019, roughly 30% of students attended a private school). The second reason is that for 6 out of 10 of individuals with oversubscribed choices, the relevant first-best versus second-best margin is two different schools, but with the same field of study. These two factors result in a usable sibling sample of roughly 6,000 observations for 2011-2019, which is far too small to be useful. In comparison, our 1977-1991 sibling sample has over 88,000 observations.

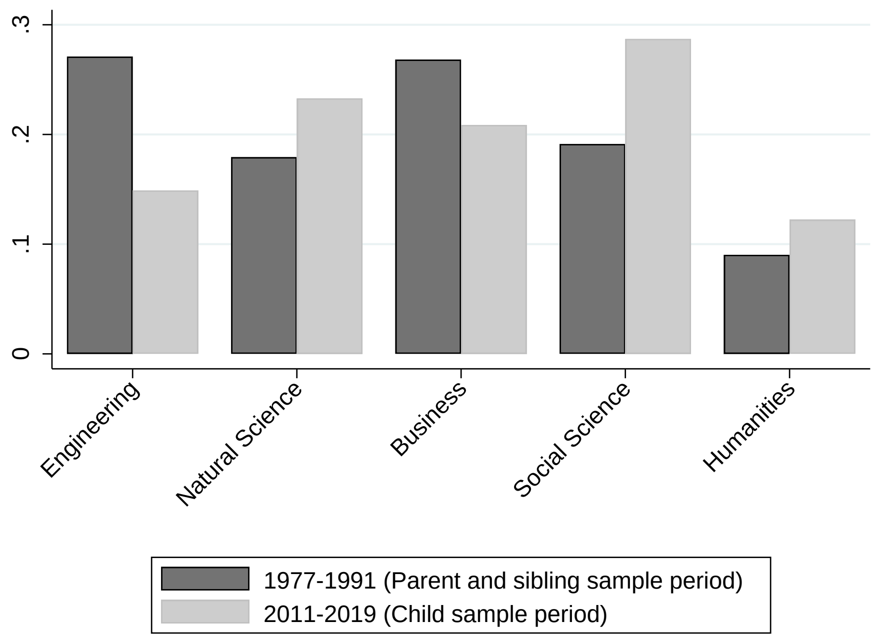
Appendix Figures and Tables for Online Publication

"Family Spillovers in Field of Study"

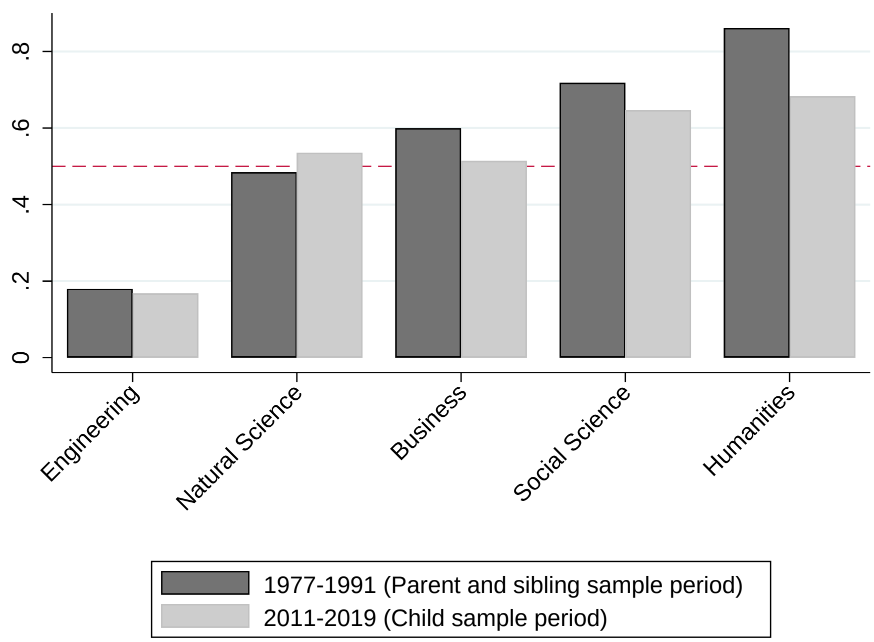
Gordon B. Dahl, Dan-Olof Rooth, and Anders Stenberg

Figure A1. Applications to academic fields of study, 1977-1991 and 2011-2019.

Panel A: Share of applicants by program

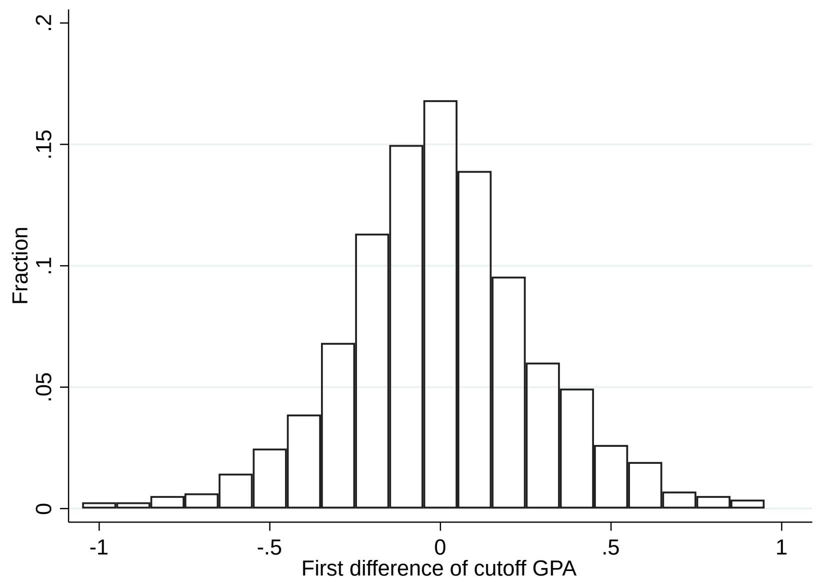


Panel B: Female share of applicants



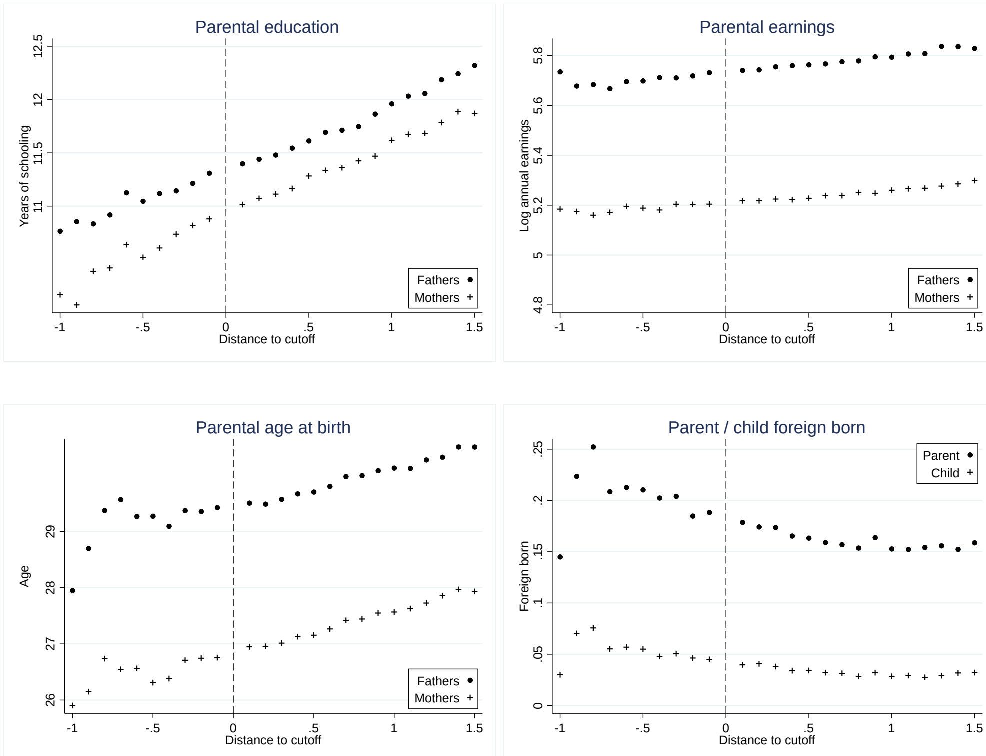
Notes: All applications to academic fields. For the years 1977-1991, N=607,767. For 2011-2019, N=558,442. The share in Humanities 2011-2019 also includes those in Arts. The dashed line marks a balanced gender composition.

Figure A2. Distribution of current minus lagged cutoff GPA, 1977-1991.



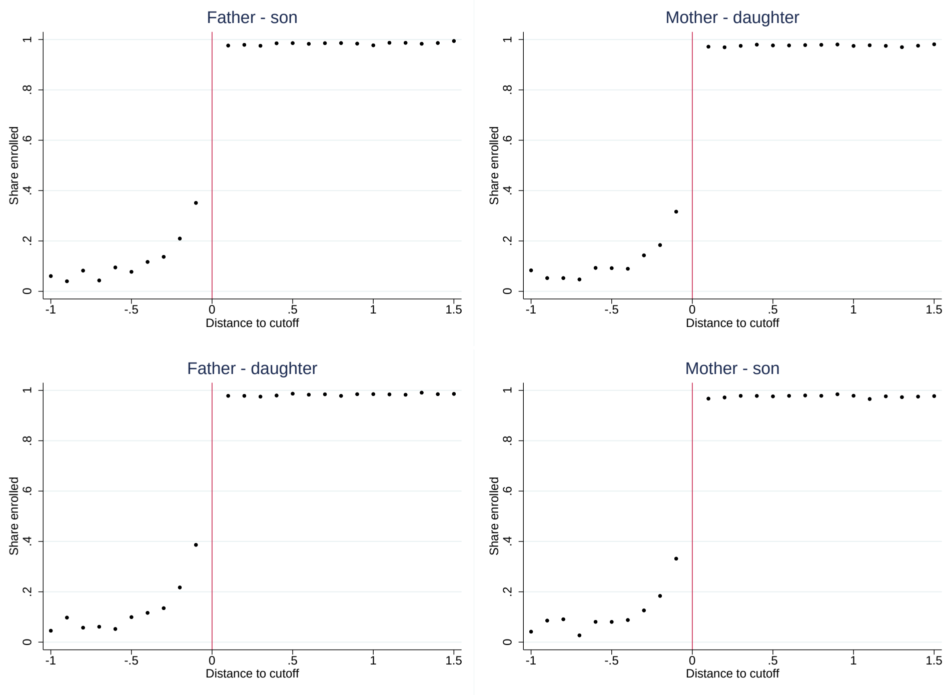
Note: Sample limited to fields of study which are competitive two years in a row in a school region.

Figure A3. Smoothness of predetermined demographic variables at the cutoff.



Notes: Each dot is the average for the relevant outcome in a 0.1 GPA bin, where GPA is measured relative to a normalized cutoff of 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).

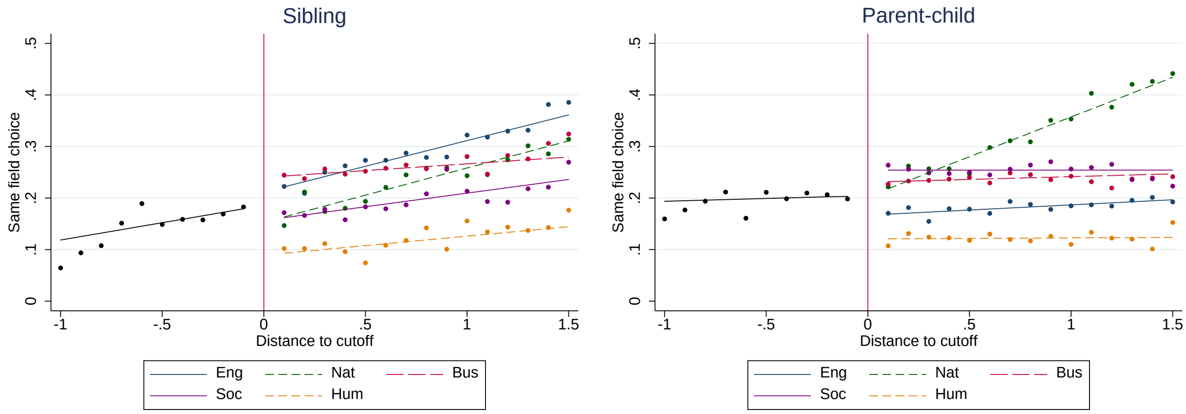
Figure A4. First stage: Share of parents enrolling in their first-choice field as a function of normalized GPA.



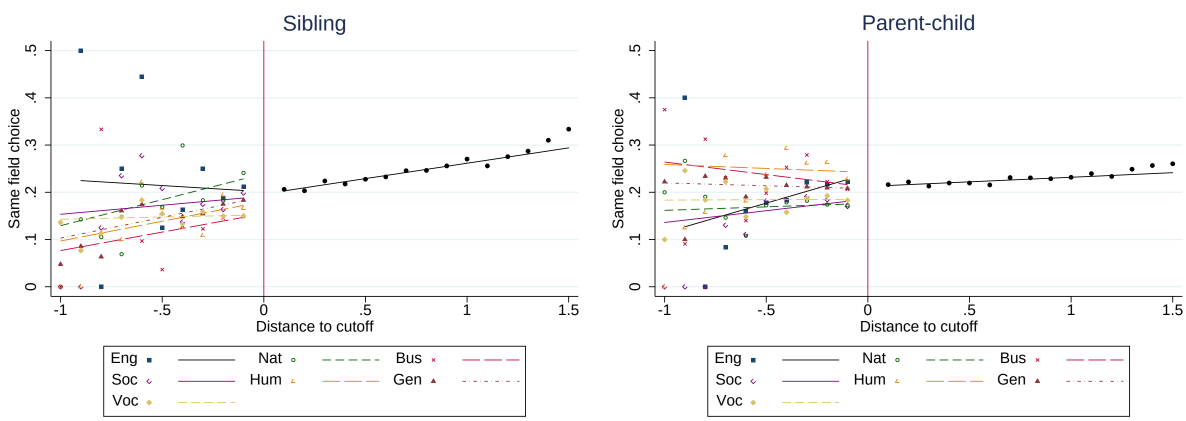
Notes: Each observation is the average share of parents who enroll in their first-best field choice as a function of their GPA. The vertical lines denote the admissions GPA cutoff (normalized to 0). See notes to Table 1 for sample sizes.

Figure A5. Comparison of the 2 versus 12-slope models.

Panel A: Single slope below the cutoff, 5 separate slopes above the cutoff



Panel B: 7 separate slopes below the cutoff, single slope above the cutoff



Notes: Panel A plots averages of the binned outcome variable for younger siblings and children against the relevant 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. Panel B 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. The graphs 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. The graphs also reveal that a 2-slope model can be too restrictive in some cases. Note that these graphs are for illustrative purposes; we never mix the 2 slope and 12-slope models in estimation.

Table A1. Siblings and intergenerational, heterogeneous effects. The probability that the younger sibling or child applies to the same secondary education program as the older sibling or parent.

Panel A: Siblings			**Panel B: Intergenerational**		
	Age gap ≤ 3 years	Age gap > 3 years		First born	Not first born
Older brother – younger brother	.021	.069***	Father – son	.035***	.021*
	(.013)	(.015)		(.011)	(.011)
Older sister – younger sister	.036**	.007	Mother – daughter	.013	-.008
	(.015)	(.017)		(.011)	(.011)
Older brother – younger sister	-.001	.017	Father – daughter	.001	.002
	(.012)	(.014)		(.010)	(.010)
Older sister – younger brother	-.043***	-.014	Mother – son	.011	.010
	(.013)	(.014)		(.010)	(.010)
N	52,357	35,942		87,405	81,798

*Notes: See notes to Table 1. Standard errors in parentheses, clustered at the family level. * p<0.10, ** p<0.05, *** p<0.01*

Table A2. Alternative measures for whether the younger sibling or child copies their older sibling or parent.

	Panel A: Siblings					Panel B: Intergenerational					
	Baseline	Same field any rank	Same field accepted	Same field enrolled	Same field completed		Baseline	Same field any rank	Same field accepted	Same field enrolled	Same field completed
(1) Reduced Form						**(1) Reduced Form**					
Older brother – younger brother	.042***	.040***	.040***	.041***	.035***	Father – son	.028***	.029***	.020***	.018**	.020**
	(.010)	(.010)	(.009)	(.009)	(.009)		(.008)	(.008)	(.007)	(.007)	(.010)
Older sister – younger sister	.023**	.025**	.029***	.028***	.027***	Mother – daughter	.003	.001	.007	.006	.006
	(.011)	(.012)	(.010)	(.010)	(.010)		(.008)	(.008)	(.008)	(.007)	(.010)
Older brother – younger sister	.007	.005	.007	.005	.003	Father – daughter	.002	.004	-.001	-.001	-.011
	(.009)	(.010)	(.009)	(.009)	(.008)		(.008)	(.008)	(.007)	(.007)	(.009)
Older sister – younger brother	-.030***	-.031***	-.030***	-.028***	-.023***	Mother – son	.012	.013	.009	.007	.001
	(.010)	(.010)	(.008)	(.008)	(.008)		(.007)	(.008)	(.007)	(.006)	(.009)
(2) IV (enrolled)						**(2) IV (enrolled)**					
Older brother – younger brother	.060***	.056***	.057***	.058***	.050***	Father – son	.040***	.042***	.029***	.027**	.027*
	(.015)	(.015)	(.014)	(.014)	(.013)		(.012)	(.012)	(.011)	(.011)	(.014)
Older sister – younger sister	.034**	.037**	.042***	.041***	.039***	Mother – daughter	.007	.005	.012	.011	.009
	(.016)	(.016)	(.015)	(.015)	(.014)		(.012)	(.012)	(.011)	(.010)	(.014)
Older brother – younger sister	.014	.010	.013	.011	.008	Father – daughter	.005	.009	.001	.000	-.014
	(.014)	(.014)	(.013)	(.013)	(.012)		(.012)	(.012)	(.011)	(.011)	(.014)
Older sister – younger brother	-.036***	-.037***	-.035***	-.033***	-.027**	Mother – son	.019*	.020*	.014	.012	.003
	(.014)	(.014)	(.012)	(.012)	(.011)		(.011)	(.012)	(.010)	(.009)	(.013)
						(3) IV (completed)					
						Father – son	.051***	.053***	.037***	.034**	.033*
							(.015)	(.016)	(.014)	(.014)	(.018)
						Mother – daughter	.011	.009	.016	.014	.011
							(.015)	(.015)	(.014)	(.013)	(.017)
						Father – daughter	.010	.014	.004	.003	-.015
							(.014)	(.015)	(.014)	(.013)	(.017)
						Mother – son	.024*	.026*	.019	.016	.005
							(.014)	(.014)	(.012)	(.012)	(.017)
N	88,299	88,299	88,299	88,299	88,299		169,203	169,203	169,203	169,203	93,559

*Notes: See notes to Table 1. The baseline outcome variable is whether a younger sibling's or child's first choice on their preference list matches the first-best field 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 field as the first-best choice of their older sibling or parent. Standard errors in parentheses, clustered at the family level. * $p<0.10$, ** $p<0.05$, *** $p<0.01$.*