Sibling Influence on College Choice

Preliminary and incomplete — please do not cite or circulate

Adam Altmejd*

September 18, 2018

Abstract

How are college choices shaped by the experiences of peers? Using registry data of applications to Swedish universities, I study how an individual's education experience influences the college applications of their siblings. Tie breaking lotteries at admission margins provide causal identification. I find that successful admission to a specific institution-program combination increases the likelihood that a sibling ranks that specific combination as their most preferred option from 2 to 3 percent. Siblings are five times as likely to follow one another to the same institution compared to the same field. The effect is stronger when both siblings are male but does not vary with the education level of the parents or with the popularity of the program. The observed spillovers seem to mainly be driven by a demand for convenience rather than the transmission of information between siblings.

*Stockholm School of Economics, adam@altmejd.se. I am grateful to my sister Nina Altmejd for wise and helpful, even though she did not follow my education choices. I thank Fredrik Lindström at UHR and Andreas Frodell at SCB for putting up with all my questions and Tore Ellingsen, Anders Stenberg, Juanna Schrøter Joensen, Erik Lindqvist, and Magnus Johanneson for valuable feedback. Financial support from the Jan Wallander and Tom Hedelius foundation is gratefully acknowledged. Any remaining errors are my own.

1 Introduction

Choosing the right college education is a daunting task. A suitable alternative needs to match one's abilities and interests and also provide good enough labor market prospects. Even though it is a choice of great consequence, applicants know very little about its impact and make it based on seemingly insufficient information (Betts, 1996; French and Oreopoulos, 2017; Hoxby and Avery, 2012; Stinebrickner and Stinebrickner, 2014).

For some people, going to college is a natural continuation of high school. Others would never even think about applying. To explain this inequality in education uptake, we need to understand both the economic and the social circumstances of schooling decisions. And what social arena is more important than the family? It is widely known that parents' education is a significant predictor of schooling (Black and Devereux, 2011), but also older siblings are influential role models (Kaczynski, 2011). Could their schooling experience have a direct impact on the choices of younger siblings?

Goodman et al. (2015) document that US siblings' college choices are strongly correlated. If an older sibling goes to college, the younger is 15-20 percentage points more likely to do so as well, controlling for family characteristics. Do siblings follow each other because they have similar preferences, or is it the education experience of one sibling that directly influences the choices of the other? The purpose of this paper is to study this potential spillover effect and to understand its underlying causes.

The many reasons why siblings follow one another can be divided into two categories. The first is related to the transmission of information. Siblings explain the application system, the available alternatives, and provide insights into the topics that they study. Information reduces uncertainty, but has an ambiguous effect on expectations. A positive signal could make the applicant more, but also less likely to apply to a specific choice, depending on prior beliefs about the program.

The second category has to do with convenience. A sibling in the same college could for example help out with housing and class assignments, provide literature and a social network. The direction of this convenience effect is less ambiguous, and likely makes the siblings follow each other.

I use tie breaking admission lotteries as natural experiments, providing exogenous variation in offers to fields and institutions. When an applicant is admitted to a program, it makes their siblings about 1 percentage point (from 2 to 3%) more likely to apply to that specific program. Interestingly, it increases the likelihood that the sibling follows to the same institution by up to 5 percentage points, but doesn't change their preferences for the field that the applicant was admitted to much. The effect is especially high for brothers, but does not vary much with the education level of the parents, nor with the quality of the program. The observed behavior is consistent with siblings following each others' education choices more by convenience than because of information transmission.

1.1 Previous Literature

This paper studies how education choice is influenced by social spillover effects between siblings. It thus bridges many strands of literature in Sociology and Economics. Here follows a short overview of the research on college choice determinants and peer effects.

To understand the importance of college education choice and what determines it, economists have recently begun to study the heterogeneous effects of college outcomes. Returns to different fields of study can vary greatly, some times with as much as the aggregate payoff of a college degree (Altonji, Arcidiacono, et al., 2016; Altonji, Blom, et al., 2012; Hastings, Neilson, and Zimmerman, 2013; Kirkebøen et al., 2016; Webber, 2014). There are also non-pecuniary benefits of attending college (Oreopoulos and Salvanes, 2011). Apart from its pure consumption value (Jacob et al., 2017), university improves marriage market outcomes (Eika et al., 2017) and health (Buckles et al., 2016).[1]

Applicants' preferences are related to their idiosyncratic earnings (Berger, 1988). Kirkebøen et al. (2016) find Norwegian students to rank programs according to their comparative advantage, but other studies show that also non-pecuniary factors play an important role (Arcidiacono, 2004; Beffy et al., 2011).

Research directly analyzing information access finds the beliefs of applicants to be quite inaccurate; about future earnings (Betts, 1996; French and Oreopoulos, 2017), costs (Hoxby and Avery, 2012), and their own field-specific ability (Stinebrickner and Stinebrickner, 2014). For example, Hastings, Neilson, Ramirez, et al. (2016) perform a large representative survey of Chilean applicants and find that they systematically overestimate the earnings of past graduates. Their respondents list prestige and accreditation as the primary reasons for their degree choice, and many do not know the average earnings in their own field of study. French and Oreopoulos (2017) argue that behavioral models are needed to explain this seemingly erratic application behavior. Based on a survey of papers that use behavioral interventions, they argue that the actions of applicants are best described by a model of inattention.

[1] The large body of research on intergenerational mobility is also relevant to understand education choice behavior. Here, correlations in both education attainment and earnings between generation is prevalent. See Black and Devereux (2011) for an overview and e.g. Fagereng et al. (2018) for some causal evidence.

There are many more papers that study the determinants of degree choice. Scott-Clayton (2012) reviews one part of this literature, noting that students underestimate costs and availability of financial aid, have noisy beliefs about future earnings, and seem to put a relatively large emphasis on non-pecuniary benefits.

The lack of information about higher education affects applicants with low socio-economic status (SES) more. In the earnings disclosure policy administered by Hastings, Neilson, and Zimmerman (2015), only low SES individuals are induced to apply for fields with higher returns. Bowen et al. (2009) find that a large proportion of highly qualified applicants from low-income households did not attend the most selective institution that they were eligible for. These applicants often forego financial aid that would be superior to what they end up receiving.

On the other hand, informing applicants about future earnings seems to have limited influence at best (Bettinger et al., 2012; Hastings, Neilson, Ramirez, et al., 2016; Pekkala Kerr et al., 2015).[2] The same is true for exposure to a field (Fricke et al., 2016). So how do individuals decide where to apply? Both face-to-face assistance with student loan application (Bettinger et al., 2012), and the provision of college students as mentors (Kosse et al., 2016) has considerable impact on the likelihood to apply to a university program. These results give credence to the theory that social transmission is an important determinant of college choice.

A substantial body of work in sociology analyzes the determinants of college choice, mostly using qualitative methods. Many studies focus on explaining low application rates among minorities. In her thesis, Kaczynski (2011) explores the sibling transmission mechanism specifically, finding multiple channels of influence. She argues that educational experience can decrease the choice set due to fear of competition, but also increase it through transmission of institution-specific knowledge and general encouragement.

The relevance of older siblings is also confirmed by Ceja (2006) and Elías McAllister (2012), who study Mexican American families with children applying to college. Ceja (2006) further emphasizes that when parents are absent or uninformed, siblings play an even more important role. Last, Butner et al. (2001) look at the choices of African-American and Hispanic women, also finding familial influence to be fundamental.

In Economics, sibling spillover effects have been studied in the context of high school choice. Schrøter Joensen and Skyt Nielsen (2017) find that when Danish students are quasi-randomly exposed to more math, their siblings apply to math programs 2-3 percentage points more often. The effect is strongest

[2] Although the low cost of these treatments can still make such interventions worthwhile.

between male sibling pairs. Dustan (2018) studies Mexico City high school applicants, who increase their likelihood of applying to a specific school by as much as 7.3 percentage points when their sibling is quasi-randomly admitted. He finds that the effect persists even when studying siblings far apart in age, where the older has already finished high school when the younger applies.

A different mechanism for sibling impact on education is presented by Brenøe (2018). She shows that the gender of the second sibling influences the choices of the first. Specifically, girls with a younger brother are 11% less likely to apply to a STEM (science, technology, engineering, and mathematics) education. Anelli and Peri (2015) study Italian students and find an effect in the same direction, with mixed-gender siblings picking more stereotypical majors.

Sibling spillovers have also been shown to matter in other contexts. Dahl et al. (2014) show that a father is more likely to take up parental leave if he has a brother who did so. Last, Konrad et al. (2002) provides evidence that children who move away from home influence their younger siblings to stay closer to the family.

1.2 Open Research Practice

The analysis presented in this paper was pre registered before the author was given access to any data (Altmejd, 2017). Such pre analysis plans are often used in experimental research to reduce the degrees of freedom a researcher has. By registering the analysis before running the experiment, the researcher can avoid the risk of finding false positives through more or less conscious data mining, and produce results that are credible (Olken, 2015; Rubin, 2007).

Ideally, a pre analysis plan should clearly map out all design choices that the researcher needs to make. While experimenters can control their data generation, empirical researchers usually guide their decisions through exploratory analysis. The problem is that such analysis will potentially bias p-values and statistical inference (Gelman and Loken, 2013). At the same time, it is very difficult to correctly anticipate all institutional obstacles and interesting sources of heterogeneity, especially in a study using population registries. While the main part of the analysis was pre registered, some supplementary tests and analyses of sub groups were not.

To make statistical inference possible also on these unregistered characteristics, I adopt a lock-box approach (Kleinberg et al., 2017). In the current version of this article, causal identification comes from tie breaking lotteries. The lock-box consists of the applicants immediately above and below the cutoff, whose admission success was arguably as good as random. The final paper will evaluate all effects also on this sample of quasi-randomized applicants. P-values from this estimation cannot be biased by post-hoc design choices.

2 Swedish Tertiary Education

University education in Sweden is publicly run and free of charge. The 40 academic institutions in our data[3] are spread out across the country and offer many different courses and programs. Some are small, with just a few hundred students enrolled, while others are large research universities. Academic credits are measured in ECTS. One semester of full time studies corresponds to 30 such credits.

University applications are centrally administered. Students are matched to programs using a strategy proof allocation mechanism — a serial dictatorship. Applications can be made to 3–5 year long programs that earn degrees, and often also directly to the individual courses that make up these programs. All options are ranked in a single ordering. It is clearly stated that it is in the students' best interests to rank choices according to their preferences. In each application period, students can be admitted to programs and courses that yield no more than 45 ECTS per semester, corresponding to one full time program and another 15 ECTS of courses.

The data includes university applications from 1993 to 2017. Starting from 2005, when universities were required to participate in the centralized application process, the data set includes all applications. For the earlier years, a small number of institutions managed admissions to some of their programs locally. Applications to these programs are not included in our data.[4] Family connections and all other control variables are provided by Statistics Sweden (SCB). Some additional data about institutional changes (in e.g. tie breaking mechanisms) for the included schools have been collected by the author.

Table 1 contains summary statistics for the Swedish population, all university applicants, as well as the subset who participate in tie breaking lotteries, constituting our sample of analysis. Approximately 6% of applicants participate in these lotteries. Since lotteries are more common in popular programs, better performing applicants from higher SES households are more likely to be included in the sample.

A few months before the start of each semester, applicants submit a ranking of their preferred programs and courses, below referred to as choices. The list can include no more than 20 such choices. Eligible applicants are ordered in a

[3] After excluding small art schools and other specialized institutions with non-standard admissions.

[4] Most of the programs with local admissions had special admission groups and would have been excluded from our analysis in any case. The only larger exception is Stockholm University, where admissions to some of the larger programs were managed locally for almost the whole period. It is unlikely that this fact has any strong bearing on our results. The results do not change much qualitatively when the sample is restricted to only include the later period. Statistical power is however greatly reduced.

Table 1. Summary Statistics

	Population	Applicants	Sample
Birth Year	1959	1980	1980
Elem. School GPA	0.01	0.46	0.60
	(0.99)	(0.79)	(0.78)
High School GPA	0.00	0.29	0.34
	(1.00)	(0.88)	(0.87)
Högskoleprovet	0.94	0.96	1.14
	(0.47)	(0.47)	(0.45)
Female	0.50	0.59	0.56
Foreign	0.23	0.19	0.11
Earnings at age 30	27.04	30.29	32.10
	(20.24)	(22.42)	(22.31)
Mother Educated	0.31	0.40	0.47
Father Educated	0.22	0.30	0.38
Siblings (N)	2.34	2.12	2.06
	(1.59)	(1.36)	(1.31)
Same gender as sib.	0.50	0.50	0.50
Unique individuals	11 145 515	2 099 773	125 442

Notes: Means and standard deviation for the whole Swedish population, all university applicants, and applicants who participate in tie breaking lotteries. Earnings are measured in $1000 and GPA scores are normalized. Score on Högskoleprovet is standardized by year with a mean of 1.

queue by their score, and ties are broken by lotteries. Whenever the choice is oversubscribed, which is true for most programs, applicants are admitted in order until all seats are filled.

To be eligible for post-secondary education, applicants must have finished high school. Certain programs have additional requirements. Engineering programs often require additional classes in science and math. Individuals who have not taken these courses in high school can supplement diplomas with preparatory courses to earn eligibility.

Universities decide what programs and courses to offer and the number of students to admit to these alternatives. These slots are divided among different admission groups (AG). An applicant can be admitted to a program in at most one group. Each AG uses a different score to rank applicants. There are three types of groups. At least a third of the applicants are admitted in groups based on GPA from high school or adult education. No less than another third are admitted on results from a standardized test similar to the SAT, called Högskoleprovet (HP). Participating in the HP group is complementary and can never decrease the likelihood of admission. A few highly selective programs in medicine use interviews for the last third (not included in the analysis), but most programs allocate also these slots by GPA.

Among all choices where the applicants are above the score threshold, they receive offers to the highest ranked 45 credits. The admission process consists of two rounds. Between the rounds, applicants can accept or reject offers, and decide if they want to stay on the waiting list for preferred choices that they have not yet been admitted to. Should they decide to wait, admissions after the second round will again include only the highest ranked 45 ECTS, even if they previously accepted offers to lower ranked alternatives.

After identifying the schools, application groups, and admissions where ties are broken by lottery, I use offers and scores to find the lottery participants. Among all eligible applicants those with scores exactly equal to the cutoff are selected. In each admission group, I identify when (a) at least one person with a score at the cutoff was admitted, while (b) one or more of these applicants were only offered a waiting list spot. In admission groups where both (a) and (b) are true, all applicants with scores at the cutoff are randomized. Only these applicants are included in the sample. An in-depth overview of the application process and the construction of the data set is given in Section A.1.

Table 2 shows means by treatment condition and results from balance tests. Since admission probability varies between experiments, and selection into different lotteries is non-random, we should not expect all pairwise comparisons to be zero. In a model predicting lottery outcomes with experiment fixed effects, a joint test of all the variables is significant ($F(15, 32881) = 9.36$, $p < 0.001$). However, when testing the same linear restriction (that all these variables are zero) but excluding the gender of the applicant, the association disappears ($F(13, 32881) = 1.55$, $p = 0.091$). A reason why the admission probability could vary by gender is given in Appendix Section A.1. All regressions include controls for the gender of the applicant.

3 Identifying the Spillover Effect

The correlation between siblings' college choices that Goodman et al. (2015) identifies has many explanations. While he controls for family characteristics, there are many unobservable features of the environment that might induce correlations in siblings' preferences. When admission is random, shared experiences cannot play any role. Studying changes in the behavior of lottery participants' siblings provides the exogenous variation that is needed.

The influence of social interactions on behavior is notoriously difficult to measure. It is hard to distinguish shared external influences from effects directly stemming from the social tie (Manski, 1993; Moffitt, 2001). Measuring

Table 2. Balance of Treatment and Control Groups

Variable	Sample		Not Admitted		Admitted		Difference
Age	22.27	(3.68)	22.32	(3.70)	22.23	(3.66)	0.09**
Age (sib)	17.63	(6.44)	17.65	(6.42)	17.61	(6.47)	0.04
Female	0.54		0.58		0.49		0.09***
Female (sib)	0.55		0.56		0.54		0.01***
Foreign	0.10		0.10		0.10		0.01**
N siblings	2.29	(1.37)	2.32	(1.39)	2.26	(1.35)	0.06***
Same gender	0.52		0.52		0.51		0.01***
Elem. school GPA	0.73	(0.76)	0.72	(0.77)	0.74	(0.74)	−0.02*
High school GPA	0.47	(0.86)	0.46	(0.87)	0.48	(0.86)	−0.02**
Högskoleprovet	1.16	(0.44)	1.15	(0.45)	1.17	(0.43)	−0.02***
Parents' Earnings	28.48	(26.27)	28.66	(26.29)	28.27	(26.25)	0.38
Mother educated	0.51		0.51		0.51		0.01
Mother field= j_{field}	0.14		0.14		0.14		0.00
Father educated	0.41		0.41		0.40		0.00
Father field= j_{field}	0.12		0.11		0.12		−0.01*
Application length	6.66	(4.07)	6.95	(4.06)	6.31	(4.06)	0.64***
Rank of j	2.57	(2.43)	3.11	(2.67)	1.94	(1.92)	1.17***
Admission prob. j	0.44	(0.25)	0.33	(0.21)	0.57	(0.23)	−0.24***
Applicants (N)	48 594		26 346		22 248		
Siblings (N)	62 571		33 977		28 594		

Notes: Standard deviations in parentheses. P-values are from t-tests of difference in means for continuous variables and proportion tests for binary variables. Only the gender of the applicant is correlates with the lottery outcome, probably because we do not perfectly capture the effect of gender-weighted lotteries used in some years. In an F-test with choice level fixed effects, the variables above the line (excluding the gender of the applicant) do not jointly explain the randomization ($F(13, 32881) = 1.55$, $p = 0.091$).

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

outcomes on the group level, it is often impossible to distinguish the effect of the group on the individual from the opposite. I circumvent this reflection problem using the exogenous variation generated by admission lotteries that influence exactly one member of the sibling pair.[5]

For a lottery participant i, randomization occurs between a preferred choice j and a lower ranked, counterfactual, choice k. For the rest of the paper, I drop i subscripts for convenience. The tie breaking lottery ensures that only chance governs whether applicant i is offered j or k, but the actual composition of j and k is their own choice. Some times, j and k will be in different fields at the same institution, and other times even the city of choice will vary. The difference between j and k can also be of much smaller consequence, for

[5] Our identification strategy furthermore allows us to estimate the endogenous peer effect — the influence from changes in peer behavior — and distinguish it from the exogenous effect, which describes the influence from varying peer characteristics (Manski, 1993). Peer-effect studies using e.g. random room mate assignments are not able to make this distinction.

example randomizing the applicant between two business programs at the same school, one with a slightly more international focus.

To estimate the spillover effect we need to know the preferred and less preferred choice of each randomized applicant. See Appendix Section A.2 for a detailed description. Starting with a data set of 2.1 million applicants submitting 65 million choices over multiple application periods, the final sample of lottery participants consists of 126 922 applicants or 247 250 sibling pairs. Only siblings who end up applying to university and submit their very first application after the lottery participant has been treated are included in the analysis. This sample consists of 63 104 sibling pairs.

We start by studying the aggregate spillover effect for each possible choice, and then continue by evaluating the effect on different sub groups. The effect is then decomposed into different sub groups for the sake of understanding the mechanisms driving siblings to follow each other.

3.1 Empirical Framework

Let z be the outcome of a lottery, with $z = 1$ when i is randomly admitted, and $z = 0$ when he or she is deferred to k. Comparing siblings of applicants with $z = 0$ to siblings of those with $z = 1$ yields the causal effect of an offer to choice j. I start by estimating a reduced form model. For each applicant i with sibling(s) $s(i)$,

$$\mathbb{I}(R_{s(i)}(j) = 1) = \alpha + \beta z + X_{s(i)}\gamma + \sum_{j} \eta_j + \varepsilon_{s(i)}.$$

Here, $\mathbb{I}(\cdot)$ is an indicator function and $R_{s(i)}(j)$ is sibling $s(i)$'s rank of choice j. The dependent variable is 1 whenever sibling $s(i)$ ranks the choice over which applicant i was randomized (j) as their highest ranked choice, and 0 otherwise. Also when j is not in $s(i)$'s ranking at all it gets the value 0. X is a set of control variables included to increase precision, and η_j are lottery level fixed effects (experiment strata).

The coefficient of interest is β. It captures the reduced form spillover effect, how much more likely the sibling is to prefer j after the applicant is randomly admitted.

Instrumentation

The sibling spillover effect is not induced by the random admission in itself, but rather from its consequences for the applicant's education. This effect can be measured directly by using the lottery as an instrument for educational attainment. We then estimate the second stage

$$\mathbb{I}(R_{s(i)}(j) = 1) = \alpha + \beta d(j) + X_{s(i)}\gamma + \sum_j \eta_j + \varepsilon_{s(i)},$$

and the first stage

$$d_i(j) = \pi_0 + \pi_1 z + X_i \psi + \sum_j \nu_j + u_i.$$

Here, the $d(j)$ is i's education attainment in choice j. I set it to 1 if i has finished at least 30 credits in j before the sibling applies. It is instrumented in the first stage by the outcome of the lottery z. Compared to the reduced form estimates above, the local average treatment effect (LATE) captured by β only measures the response of siblings to complying applicants. It will thus measure the difference between those who take at least 30 ECTS in j and have $z = 1$ and those with no credits in j and a $z = 0$.

The assumptions underlying IV estimation need to hold for β to not be biased and correctly reflect the LATE. A valid instrument needs to satisfy the exclusion restriction. Since the only plausible channel through which the outcome of an admission lottery could impact sibling preferences is through the education itself, there is no reason why the assumption would not hold. Lottery participants also need to actually follow the outcome often enough, or the instruments will be weak. In Appendix Table A4 we see large enough first stage estimates, with F-statistics far above 10.

IV can be interpreted as a LATE as long as the monotonicity assumption holds. An activated instrument should not move applicants into not taking the treatment (no defiers). Since j is preferred to k it is unlikely that winning the lottery would induce applicants to study k instead. Since less preferred applications are automatically withdrawn, it is often impossible for a lottery winner to study their less preferred choice the same year. Instead, they would have to first reject the offer and then reapply to k in the following year. The existence of defiers therefore seems very unlikely.

Controlling for Selection with Lottery Fixed Effects

Our data set consists of a large number of experiments where the applicant is either admitted to j or deferred to k. Selection into these experiments is non-random. Applicants choose themselves which programs and institutions to apply to. They can also chose to write the HP to participate in another admission group.

Is there a risk of confounding the correlated preferences of siblings with spillover effects? The aggregate estimation is based on averages over multiple experiments where assignment probabilities vary. If correlated preferences are more common in lotteries for choices where the admission probability is high there is indeed such a risk. The following extreme example is an illustration.

Say our data set was made up of two lotteries over preferred choices. One for an engineering program and another for a medical program. Both programs have ten applicants who participate in the lotteries, but only one engineer receives an offer while 9 medical school applicants do. In other words, the probability of winning is much higher for the medical school applicants. At the same time, all of their siblings also chose to become doctors, while the engineering siblings have no such family pressure and choose a variety of subjects. If one pools these observations in a regression, it will look like almost all lottery winners had siblings who followed them, while there was no systematic preference for the preferred choice among the losers. In reality, what one captures is that siblings have more correlated preferences for choices with higher admission probabilities.

For the situation described in the above example to be problematic to the analysis presented here, both the degree to which choices are correlated and the admission probabilities need to vary between programs. It is far from impossible that such systematic variation exists in our data. All specifications include experiment level fixed effects (η_j) to get around these problems. In a model with experiment fixed effects, the measured effect cannot be influenced by any differences between experiments.

Each choice j is a separate lottery. To interpret these lotteries as experiments we also need to define the relevant control groups. In our case, it is the next-best alternative k that constitutes this control group. To give an example, the kind of people who end up in a lottery for medical school but with an engineering program as fallback could be systematically different from those with a business program as their k. Kirkebøen et al. (2016) argues that economic returns need to be estimated separately for each combination of preferred and less preferred choice. Due to the limited sample size, the pre-analysis plan does not stipulate the inclusion of fixed effects for next-best choices. However, Table A7 shows that it does not change estimates much.

4 Results

When an applicant is randomly admitted to a choice, the likelihood that their siblings will follow and prioritize that choice in their ranking increases by almost 1 percentage point, from 2.1 to 2.9 percent. In Table 3 we see this reduced form effect, and that it is stable over all specifications. Furthermore, the last column shows the results from an IV regression, where random admission is used as an instrument for if the applicant finishes at least 30 ECTS before the sibling applies. The LATE is more than twice as high as the reduced form effect, likely due to the large share of always takers, decreasing the first stage.

Table 3. Sibling Choice Spillovers

	Outcome: $\mathbb{I}$ (Sibling ranks j as top choice)			
	(1)	(2)	(3)	(4)
Admitted j	0.009^{***}	0.008^{***}	0.009^{***}	
	(0.001)	(0.001)	(0.001)	
>30 Choice				0.025^{***}
				(0.004)
Choice f.e.		$\times$	$\times$	$\times$
Controls			$\times$	$\times$
Control $\bar{y}$	0.017	0.017	0.017	0.017
$\mathbb{R}^2$	0.001	0.112	0.116	0.117
Observations	62 801	59 843	59 823	59 823

Notes: The dependent variable is binary and equal to 1 whenever the sibling ranks j as their top choice. The last column is an IV regression where the lottery outcome is used as an instrument for if the applicant completes at least 30 ECTS before the sibling applies. It has a first stage F-statistic of 314.8. Choice fixed effects: year $\times$ choice $\times$ ag dummies. Controls: age, gender, parental education level and income, and if parents' education is in field j. Standard errors in parentheses.
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

Contrary to our initial pre registered hypothesis, the table shows a positive and stable spillover effect between siblings. Is the effect of economic significance? On average only 2% of siblings prefer the exact same choice, meaning that even this tiny effect translates to an increase of almost 40%. The reason our baseline is so low is that choices are very narrowly defined. The sibling needs to apply to the exact same program at the exact same institution for the outcome variable to be 1. When we look at coarser measures, like if siblings go to the same school (Table 4), the absolute size of the estimates increases to levels more similar to those found by other researchers (Dustan, 2018; Schrøter Joensen and Skyt Nielsen, 2017).

The pre registered hypothesis states that underlying heterogeneity would make the aggregate spillover effect indistinguishable from zero, or perhaps weakly positive. Like Kaczynski (2011) finds, I argued that some siblings would follow, while others would avoid the specific choices of their brothers and sisters to steer clear of any competition.

I predicted a noisy null effect for the main analysis, since if information transmission was a key driver of sibling spillovers, the direction of the effect would be ambiguous. Obtaining information makes the younger sibling update already held beliefs. If they think really highly of a program, even a positive signal could lead to anti-following when expectations are overoptimistic.

While we see a main effect in Table 3 that is clearly positive and very stable, contrary to the hypothesis, it is small. To understand what mechanisms drive this positive aggregate effect, and see if it might mask anti-following or other group-specific patterns, we study a number of subsets of the sample below.

To study what exactly it is that drives this imitation effect, we now turn to the many heterogeneous effects in our data, making use of the different experiments created by the variation in admission margins.

4.1 Following to Fields, Institutions and Cities

In Table 4, the three rightmost columns show estimates of the effect on three different sub samples, where either field, institution, or city varies between j and k. In each case the outcome variable is redefined to be equal to 1 as long as the sibling ranks the field, institution, or city of j as their most preferred choice.

The reason why the main analysis is at the most granular level is that it permits the use of the whole sample. Every randomization moves the applicant between two different choices. If we instead group the different alternatives by e.g. college field or institution like in Table 4, many margins will have to be excluded, since they no longer move the applicant between different categories. This it the reason why sample sizes vary between the columns in the table. They overlap because many lotteries move applicants between both two different institutions, and two different fields.

Table 4. Spillovers by Field, Institution, and City

	Sibling follows to same:			
	(1)	(2)	(3)	(4)
	Choice	Field	Institution	City
Admitted j	0.009***	0.008	0.045***	0.049***
	(0.001)	(0.005)	(0.005)	(0.005)
Choice f.e.	×	×	×	×
Controls	×	×	×	×
Control $\bar{y}$	0.017	0.160	0.167	0.237
$\mathbb{R}^2$	0.116	0.209	0.212	0.240
Observations	59 823	36 824	46 768	44 835

Notes: Choice fixed effects: year × choice × ag dummies. Controls: age, gender, parental education level and income, and if parents' education is in field j. Standard errors in parentheses.
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

We see a clear pattern in Table 4. The aggregate imitation effect seems to be driven by the tendency of siblings to prefer the same school.[6] When the lottery drives the applicant into a specific field, on the other hand, the sibling's preferences for that field does not change much at all.

Note that the baseline differs between columns. Naturally, a larger fraction of siblings want to study in the same city (0.237) compared to the number who prefer the exact same choice (0.017). The relative effect is therefore smaller, with the 5 percentage point increase corresponding to about 20%. Columns (2) and (3) have almost identical baselines, meaning that about the same number of control group siblings prefer the field or institution of j. The difference in treatment effects means that it is five times as likely that a sibling ends up following the applicant to the same institution. The corresponding difference between IV-estimated LATEs is a factor of three, as we see in Appendix Table A3.

If information transmission was an important part of the spillover effect, the field estimates should be stronger. After all, relative to the other two columns, the difference in information transmitted between siblings is arguably the largest when randomization happens between two fields. But instead, we see a fairly strong effect when the treatment randomly allocates the applicant between two different cities. One possible explanation for these results is that, rather than due to information transmission, siblings follow each other because it is convenient and to minimize costs. Moving to a new city could be less cumbersome if one family member is already present, for example.

4.2 Sibling Gender

Just like previous research on sibling spillovers in STEM subjects (Schrøter Joensen and Skyt Nielsen, 2017), Table 5 shows larger effects when both siblings are male. The absolute effect is almost twice as large as in other groups, but male pairs also have a higher baseline. In other words, male siblings have more correlated preferences, but they also seem to follow each other more often. Previous research has stressed that siblings might not want to pick the same fields to avoid direct competition. The fact that males are more competitive could partly explain the observed differences, but experimental evidence on gender differences in preferences for competition does not show any specifically large effect for competition in male-only groups (Booth and Nolen, 2012).

[6] Since most Swedish cities only have one university, the difference between the estimates of the two rightmost columns of Table 4 mostly comes from the relatively large number of schools in Stockholm

Table 5. Choice Spillovers by Gender

	Applicant/Sibling			
	(1)	(2)	(3)	(4)
	M/M	M/F	F/M	F/F
Admitted j	0.019^{***}	0.005	0.008^{*}	0.008^{*}
	(0.005)	(0.003)	(0.003)	(0.004)
Choice f.e.	×	×	×	×
Controls	×	×	×	×
Control $\bar{y}$	0.025	0.014	0.011	0.022
$\mathbb{R}^2$	0.207	0.230	0.213	0.219
Observations	10 859	12 871	11 818	16 054

Notes: Choice fixed effects: year × choice × ag dummies. Controls: age, gender, parental education level and income, and if parents' education is in field j. Standard errors in parentheses.
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

4.3 Parental Education Level

Sociologist argue that siblings and other relatives with college experience are more important for college decisions in families with lower socio-economic status (Ceja, 2006). When neither parent has any university education, the randomly admitted applicant is the sibling's only source of information about college, potentially increasing the strength of the treatment effect. I test this claim by studying sibling influence within families where the parents have different levels of education.

Table 6. Choice Spillovers by SES

	Number of Educated Parents		
	0	1	2
Admitted j	0.009^{***}	0.008^{*}	0.013^{***}
	(0.003)	(0.003)	(0.004)
Choice f.e.	×	×	×
Controls	×	×	×
Control $\bar{y}$	0.016	0.015	0.023
$\mathbb{R}^2$	0.213	0.204	0.170
Observations	20 327	16 978	17 259

Notes: Choice fixed effects: year × choice × ag dummies. Controls: age, gender, parental education level and income, and if parents' education is in field j. Standard errors in parentheses.
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

Surprisingly, Table 6 shows an ambiguous effect, with the highest degree of following behavior among siblings with two educated parents. While the differences between columns are small, the rightmost group has more college enrollment overall, but also a higher baseline. Since the treatment is between different choices, it is not possible to measure the spillover effect on the extensive margin, from going to university or not. It is possible that the stronger effect of familial influence discussed in e.g. Ceja (2006) mainly shows up on this margin.

The Appendix contains a number of additional sub group analyses that were also pre registered but do not provide much additional information. Random admission of a sibling has no differential effect on following depending on if the program is more competetive (Table A1), nor does successful admission improve the academic performance of the younger sibling (Table A2).

4.4 Robustness

Table 7. Placebo Tests

	Sibling follows to same:			
	(1)	(2)	(3)	(4)
	Choice	Field	Institution	City
Admitted j	0.006^{***}	-0.006	0.008	0.012^{*}
	(0.002)	(0.006)	(0.005)	(0.006)
Choice f.e.	×	×	×	×
Controls	×	×	×	×
Control $\bar{y}$	0.018	0.162	0.161	0.232
$\mathbb{R}^2$	0.160	0.239	0.260	0.291
Observations	52 898	31 269	40 106	37 681

Notes: The regressions in the table only include those who applied before their sibling was randomly given an offer. Choice fixed effects: year × choice × ag dummies. Controls: age, gender, parental education level and income, and if parents' education is in field j. Standard errors in parentheses.
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

If the effect picked up is properly identified and not a residue of the correlated preferences between siblings, those applying before the randomization takes place should not be affected by the outcome of the lottery. However, if what is measured is correlated preferences, a seemingly non-chronological effect could show up. We study this placebo test in Table 7, where the same regressions as in Table 3 are run, but now on siblings that apply before the randomization happens.

Some of the coefficients are positive, but smaller and noisier than before.

Measurement errors might explain the positive estimates. Especially in the older data, detailed information about the use of lottery tie breakers is usually not available. It is possible that certain admission margins were incorrectly specified, and e.g. gender quotas were used without it being registered. However, interacting the fixed effects with gender does not remove the placebo effect completely (see Table A9), indicating that misspecification alone cannot explain the pattern.

5 Conclusions

I find substantial and stable spillover effects from admissions, mostly driven by siblings following each other to the same schools rather than into the same field of study. The effect is twice as strong when both siblings are male, but does not differ with the education level of the parents, nor with the competitiveness of the program.

It is not strange that preferences for fields of study are much less influenced by the education experience of siblings. Applicants might already have decided what career they are aiming for at a young age. Coming from different high school tracks might even prohibit some siblings from following into the same field. Siblings may also avoid going into the same field to not compete too much as Kaczynski (2011) posits. Studying the same topic makes it much easier to compare not only grades, but also career success later in life. Since the variation in earnings between fields is larger than that between institutions, a change in field preferences has a larger impact on earnings. Choosing to go to the same institution, on the other hand, matters little for earnings but still provides many benefits, like access to a social network and potentially housing.

The analysis in this paper confirms previous work (Goodman et al., 2015). Just like in the US, about a fifth of all Swedish siblings apply to the same university. Admission of an applicant increases the likelihood that their sibling will apply to the same school by about 5 percentage points. In other words, about a fourth of the correlation between siblings in university choice can be ascribed to a causal spillover mechanism.

Our findings also give some credence to the multiple and sometimes contradicting effects described by Kaczynski (2011). She finds that applicants avoid the choices their siblings have made due to fear of rejection and competition. However, she also argues that having older siblings attending university made the applicant more likely to enroll in college, and more interested in the specific institution of their sibling. Our results partly support her conclusions. The weak influence on field preferences could be due to fear of competition. But on the aggregate, the will to follow a sibling dominates any such worries.

The previous quantitative studies on sibling spillovers in education has only looked at high school choices. We see weak effects for field spillovers compared to the fairly strong influence on high school STEM course take up found by Schrøter Joensen and Skyt Nielsen (2017). The substantial effect on institution preferences are in line with Dustan (2018). The effect he finds is stronger, however, which could potentially be explained by the fact that parents have more control over high school choice.

To summarize, we find positive spillover effects between siblings, mostly driven by changes in preferences for location rather than field of study. What are the implications of these findings? For one thing, universities should be aware of the positive spillover effects from admissions. When trying to increase enrollment of students from minorities, through e.g. affirmative action, familial spillovers will increase the effectiveness of policies. Second, it is interesting that sibling preferences for fields are only weakly affected. If information transmission was a key mechanism explaining correlated choices, one would expect a positive spillover effect between fields, and from higher quality choices. Instead, it seems that siblings follow each other out of convenience.

A Appendix

A.1 The Admission Process

This section provides an in-depth description of the university application process, selection of applicants, the use of tie-breaking lotteries, and a description of how the sample was constructed.

The slots that a university allocates to a specific course or program are divided into admission groups (AG). We only include the three main groups, with scores from three sources; GPA, HP, and a special kind of adult education called Folkhögskola (BF). At least a third of the slots need to be used to admit applicants from the GPA and BF groups, while another third uses the HP group. The last third can be allocated however the institution prefers, but is usually also directed to the GPA group.

There is a large number of different GPA admission groups, partly because the Swedish high school grading system has changed multiple times during the sample period. The respective size of each group is proportional to the relative number of applicants within each such subgroup. If there are twice as many applicants with grades from an older system, their admission group will be twice as large as the one for students with grades from the new system.

The groups can be divided into two kinds; those that use the applicant's raw high school GPA, where no results from adult education can be included, and those where supplemented scores are allowed. Naturally, since grades from adult education can be improved (multiple times if needed) groups in the second category usually have much higher admission cutoffs.

In contrast, the results from Högskoleprovet are normalized year by year, with scores ranging between 0 and 2. The score was reported with one decimal (20 different levels) until 2012, when also half decimals (increasing to 40 levels) were allowed. This coarse score distribution makes tie breaking quite common.

Last, a small number of applicants are admitted in a special GPA group with grades from a Folkhögskola (BF in Figure A1). Folkhögskola is a kind of adult education, mainly for those who did not finish high school. These schools use a different GPA system than the regular high schools, and since there often are very few slots for students with such grades, admission thresholds are usually high.

During the application process, each choice has a separate ordering of applicants for every admission group. Some individuals are present in several orderings, while others only have one valid score. Applicants are admitted one by one, always from the admission group that has the most slots to fill. If an applicant is admitted in one group, he or she is removed from the queue in all other groups.

It is only when all slots have been filled that the individual's own ranking is taken into account. Offers are given to the first 45 ECTS that an applicant has been successfully admitted to. Any lower prioritized application is withdrawn. The allocation process is then repeated to fill spots with applicants further down the queue, and successful admissions over 45 credits are yet again removed. This process continues until all choices have been filled. Students that still have not received any offers are put on waiting lists for the second admission round.

Applicants can view their offers at the end of the first round. For choices to which the applicant was not admitted but that are preferred to any received offers, the position on the waiting list is shown. Applicants must then choose to accept and/or reject each offer, and decide if they want to stay on waiting lists for a chance to receive an offer to prioritized choices at a later stage in the process.

The second round works just like the first. Applicants are only admitted to their highest ranked 45 credits. This means that opting to stay on the waiting list at the end of the first round could mean that a lower prioritized but previously accepted offer is withdrawn. If the applicant's preferences have somehow changed, for example because of an initiated move to a specific city, he or she should make sure not to stay on waiting lists for choices that are no longer preferred.

At the end of the second round, applicants are yet again offered the possibility to stay on waiting lists to preferred choices that they were not admitted to. At this point, the lists are sent to the local institutions who use them to supplement for potential drop outs. An offer from the waiting list after round two no longer means automatic withdrawal from less preferred alternatives. Applicants choosing to stay on the waiting list after round two can receive offers a few weeks into the semester. Since most schools anticipate some degree of attrition and recruit extra students from the start, supplementary offers to applicants on the waiting lists are quite rare.

Local admissions off the waiting lists are not registered in our data. Our natural experiment is about randomly allocated offers and should be seen as a measure of the intention to treat-effect. Some of those that do not win the lottery will end up receiving local offers. To estimate the treatment effect of actually having a sibling that studies the program they preferred, rather than their next-best alternative, instrumental variables estimation is required.

Tie Breakers

When eligible applicants to oversubscribed choices are ranked by score, ties need to be broken. During most admissions in our data, lotteries are used to do so. Our sample consists of applicants who participate in lotteries for their preferred choice. Figure A1 is a histogram over the distribution of randomization cutoffs in the different admission groups and their frequency.

In HP admission groups, lotteries occur all throughout the performance distribution. However, randomization in GPA-based cutoffs is mostly occurring in cases where applicants have really good scores. The reason for this pattern is twofold. For one thing, HP scores are a lot more coarse than GPA, making ties more common over the whole distribution. Moreover, HP results are normalized while high school grades are awarded for absolute performance levels.

The law regulating admission changed multiple times during our sample period and so did the rules on how to break ties. Complementary to lotteries, a number of other options have been allowed during the sample years (see changes in 4. kap 12 § *Högskoleförordningen* 1993). The current law allows for tests (usually HP) and interviews, but during 2006–2010 schools were also allowed to prioritize tied applicants by gender.

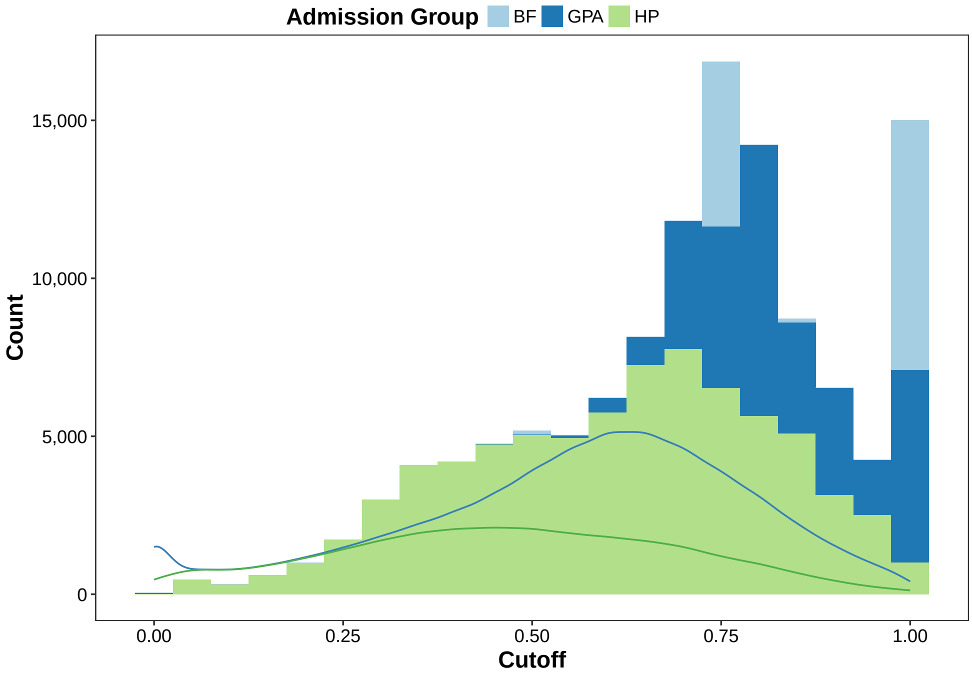
For a few years in the early 2000s, universities admitted men and women separately. Gender quotas were deemed unconstitutional in 2006, but universities continued using gender to break ties. In practice, gender weights were applied to lotteries, to give higher admission probability to members of the underrepresented gender. These weighted lotteries were used in most admissions between 2000–2008 (Lövtrup, 2008). Between 2008 and 2010, it was instead ensured that whenever ties needed to be broken lotteries were done separately by gender, and men and women were admitted in alternating order. Under both these regimes, all tied applicants are randomized, but admission probabilities change with gender. Starting in spring 2011, the use of gender to break ties was prohibited. All specifications use controls for gender, but Table A8 instead includes gender-interacted fixed effects, without any noticeable impact on results.

In certain admission groups, the most popular programs are sometimes oversubscribed by applicants with perfect GPA scores, meaning that everyone is tied. If a lottery tie breaker is used, all admitted students will have been randomized, which is why we see so much mass in the rightmost column of Figure A1. To avoid this, many medical programs have recently started using HP scores to break ties. HP tie breakers have been introduced gradually, but are now used for all medical programs in Sweden. A few universities also use HP tie breakers for other admissions. Lotteries are then still employed in the rare cases when both GPA and HP scores coincide.

Strategy Proofness

It is a well-known fact that the serial dictatorship is a strategy proof and pareto efficient allocation mechanism (Svensson, 1999). The Swedish system is only

Figure A1. Distribution of Cutoffs



Notes: Histogram over distribution of cutoffs where randomization happens. Scores are standardized so that 1 equals the top score in each admission group (e.g. 2.0 in the HP group) and 0 the lowest. It is clear that most randomization in the GPA-based admission groups comes from lotteries at the top of the distribution. HP randomization is more spread out according to the actual distribution of HP scores. Overlaying the histogram are two solid lines, density plots of the actual grade distributions in the HP and GPA groups respectively. While the HP sample represents the population quite accurately, the applicants that are randomized in GPA groups are predominantly top performers.

slightly different from the classical setup. While likely not pareto efficient, it is indeed rational for applicants to reveal their true preferences when applying.

The main difference from a classical setup is that the application process is divided into two rounds and that there are multiple admission groups. The choice set in the second round is limited by the ranking of the first. In the second round, the applicant choses to accept or reject any potential offers, and to stay on or withdraw from the waiting list for each preferred choice. Since the algorithm is the same, the only reason to take any action is if preferences have changed in some way. The second round ranking should thus also reflect the true preferences of the applicant.

A.2 Constructing the Sample

The data used in this study comes from three sources. University applications are currently managed by Universitets- och Högskolerådet (UHR) who provided application data for the years 2008–2016. The now closed Verket för Högskoleservice (VHS) administered the application system before that. Applications from 1993–2005 were retrieved from the Swedish National Archives. The last data source contains individual characteristics of applicants, parents, and siblings, and was delivered by Statistics Sweden (SCB).

After joining the data sets and identifying lotteries, pairs of lottery choice and counterfactual alternative can be constructed. But the actual lotteries are not between the two options j and k. Rather, winners are admitted, while losers are deferred to the next choice in their ranking, where possibly a second lottery could await them. Here, I describe exactly how these lotteries are identified and their respective counterfactual outcomes.

I first remove applications where the applicant is not eligible and identify all applicants that participate in lotteries. A lottery participant is a person who has a score in at least one admission group exactly at the cutoff, and no score above the cutoff in any other group. There must also be other applicants in that group with scores exactly at the cutoff, and at least one of them must have received an offer.

An applicant's highest prioritized lottery is a candidate for j. If they win, I assign said choice to j and the next choice in their ranking that they would have been admitted to as k. I also set $z = 1$.

In cases when the lottery is lost I instead look at the choice to which the candidate was deferred. Should they be successfully admitted to this choice, no matter if through a lottery or not, I use the same assignment as before but set $z = 0$. If this second-best option constitutes an unsuccessful lottery, I declare a different j/k margin. In that case, I use the second option as j (with $z = 0$) and assign as k a less preferred choice that yielded an offer. After finding all pairs of preferred choice j with a lottery and counterfactual outcome k, I then use the j/k margin where j has the highest rank. At this point we have one margin per applicant such that (a) there is a lottery over j and (b) if the lottery fails the applicant is admitted to k instead.

Some applicants really want to get in to a specific program and will do everything they can to achieve their goal. Writing the HP multiple times seems to improve results slightly, and applying for multiple consecutive semesters could make admission more likely. Under such circumstances, a failed lottery is a pretty strong signal that a successful admission is possible and that the student should indeed re-apply. To avoid any issues of endogeneity through selection caused by behavioral responses from re-applications I only look at the first time an applicant participates in a lottery for a degree program.

At this point, I match the lottery participant to the applications made by their siblings. In some cases, one sibling can have multiple lottery participants to join to. I then only allow the sibling to be influenced by the latest lottery. Our final sample consists of exactly one application per sibling, but potentially multiple occurrences of a single lottery, should that applicant have several siblings. For this reason, all regressions have standard errors clustered at the level of the lottery participant.

A.3 Quality Interaction

One of the main hypothesis in our pre analysis plan was that even though the aggregate results would be too noisy to identify any following behavior, we should see spillovers for the most popular programs. The reason is that among programs of high quality, the information signal sent by the lottery participant should be more positive, increasing the likelihood that the sibling updates their beliefs upward. We present three different measures of popularity interacted with the treatment assignment variable in Table A1, neither which have any impact on spillovers.

A.4 Impact on Siblings' Grades

A second avenue of study pre specified in the analysis plan is the impact of admission on the academic performance of the sibling. I wanted to test if the success of the older sibling motivates the younger. Any effect was hypothesized to be stronger when the difference between j and k in terms of difficulty and competitiveness is large. In Table A2 we see a regression of the individual difference between standardized elementary school GPA and high school GPA on the outcome of the lottery. We only look at siblings who started but have yet to finish high school when the lottery happens. In the last column we instead employ a between-subjects design, using only the standardized high school GPA as dependent variable. It doesn't seem like the education experience of the older sibling has any meaningful impact on the grades of the younger, not even when the difference between j and k is larger (the two middle columns).

A.5 More IV Estimates

Table A3 shows results from IV models where we look at the effect of the applicant at least taking 30 ECTS in the field, institution, or city to which he or she is randomly admitted. As with the aggregate effect in Table 3, the causal effect on the complier group is substantially larger than the reduced form. The first stage is presented in Table A4.

Table A1. Choice Spillovers and Quality Interactions

	Outcome: $\mathbb{I}$(Sibling ranks j as top choice)			
	(1)	(2)	(3)	(4)
Admitted j	0.009^{***}	0.009^{***}	0.006^{*}	0.009^{***}
	(0.001)	(0.002)	(0.002)	(0.002)
Popularity		0.000		
		(0.000)		
z × Popularity		0.000		
		(0.000)		
Pop. Rank			0.000	
			(0.000)	
z × Pop. Rank			0.000	
			(0.000)	
Cutoff Rank				0.000
				(0.000)
z × Cutoff Rank				0.000
				(0.000)
Choice f.e.	×	×	×	×
Controls	×	×	×	×
Control $\bar{y}$	0.017	0.017	0.017	0.017
$\mathbb{R}^2$	0.116	0.117	0.117	0.117
Observations	59 823	58 479	58 631	58 631

Notes: All interaction effects are 0 down to the fifth decimal. Choice fixed effects: year × choice × ag dummies. Controls: age, gender, parental education level and income, and if parents' education is in field j. Standard errors in parentheses.
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

A.6 Robustness

A possible criticism is my choice of dependent variable. I decided to study the highest ranked program for two reasons. First, compared to a continuous variable based on the actual ranking, a binary variable makes it possible to include siblings that do not rank the lottery choice at all. Second, requiring that $R_{s(i)}(j)$ is 1 makes the meaning of an activated indicator variable as unambiguous as possible. A sibling including the choice somewhere else in their ranking could have higher prioritized alternatives that are easier to get in to, and does so without actually believing that they would ever study said choice.

Table A5 shows estimates from the choice-level spillover effect when the outcome variable includes more ranks. The second column uses an indicator for if j is in the siblings' top three, the third is activated as long as the sibling includes

Table A2. Impact of Admission on Sibling Performance

	ΔGPA			HS GPA
	(1)	(2)	(3)	(4)
Admitted j	0.022			0.021
	(0.013)			(0.015)
Sib. ES GPA	-0.348^{***}	-0.349^{***}	-0.359^{***}	
	(0.014)	(0.023)	(0.020)	
z × Cutoff Diff		−0.010		
		(0.021)		
z × Pop. Diff			0.000	
			(0.000)	
Choice f.e.	×	×	×	×
Controls	×	×	×	×
Control $\bar{y}$	−0.232	−0.220	−0.235	0.274
$\mathbb{R}^2$	0.300	0.350	0.342	0.334
Observations	24 103	9735	12 734	24 103

Notes: All interaction effects are 0 down to the fifth decimal. Choice fixed effects: year × choice × ag dummies. Controls: age, gender, parental education level and income, and if parents' education is in field j. Standard errors in parentheses.
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

the choice somewhere in their ranking. The last column uses the actual rank of j in the sibling's application. The regression is only run on those siblings who actually have j somewhere in their ranking. It is measured relative the the length of the ranking, with 1 meaning highest ranked. The baseline means that siblings rank j on approximately position two out of five, but slightly higher should the lottery participant be admitted. It seems thus that most of the effect we find comes from the extensive margin, i.e. that more siblings include j in their ranking when $z = 1$.

Each applicant can be subject to multiple lotteries. Following the analysis plan I only use the highest prioritized margin that includes an admission. But in some cases the applicant fails many lotteries before being admitted, and we could include them in the control group to increase power. Table A6 does so and shows very similar but somewhat larger estimates compared to the main specification.

If selection into next-best choices is somehow related to the sibling connection, our estimates could be slightly wrong since our fixed effects are only at the preferred choice margin. In Table A7 we include fixed effects for each combination of j and k. Results are slightly smaller, but do not change much. Because of the large number of fixed effects, they should be interpreted with caution.

Table A3. IV Regressions

	Sibling follows to same:			
	(1)	(2)	(3)	(4)
	Choice	Field	Institution	City
>30 Choice	0.025***			
	(0.004)			
>30 Field		0.052		
		(0.034)		
>30 Inst.			0.150***	
			(0.015)	
>30 City				0.170***
				(0.018)
Choice f.e.	×	×	×	×
Controls	×	×	×	×
Control $\bar{y}$	0.017	0.160	0.167	0.237
$\mathbb{R}^2$	0.117	0.207	0.229	0.263
Observations	59 823	36 824	46 768	44 835

Notes: Choice fixed effects: year × choice × ag dummies. Controls: age, gender, parental education level and income, and if parents' education is in field j. Standard errors in parentheses.
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

Table A4. First Stage

	Outcome: >30 ECTS in j before sibling applies,			
	(1)	(2)	(3)	(4)
	Choice	Field	Institution	City
Admitted j	0.339***	0.147***	0.299***	0.288***
	(0.005)	(0.007)	(0.006)	(0.006)
Choice f.e.	×	×	×	×
Controls	×	×	×	×
F	312.401	41.655	183.347	161.681
$\mathbb{R}^2$	0.391	0.342	0.366	0.352
Observations	59 823	36 824	46 768	44 835

Notes: First stage estimates of Table A3
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

Table A5. Alternative outcome variables

	j among sibling's			
	(1) Top 1	(2) Top 3	(3) All	(4) Rank of j
Admitted j	0.009***	0.011***	0.011***	0.039*
	(0.001)	(0.002)	(0.003)	(0.015)
Choice f.e.	×	×	×	×
Controls	×	×	×	×
Control $\bar{y}$	0.017	0.027	0.073	0.718
$\mathbb{R}^2$	0.116	0.130	0.153	0.356
Observations	59 823	59 823	59 823	2318

Notes: Choice fixed effects: year × choice × ag dummies. Controls: age, gender, parental education level and income, and if parents' education is in field j. Standard errors in parentheses.
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

Table A6. Including multiple lotteries per applicant

	Sibling follows to same:			
	(1) Choice	(2) Field	(3) Institution	(4) City
Admitted j	0.009***	0.007	0.049***	0.054***
	(0.001)	(0.005)	(0.004)	(0.005)
Choice f.e.	×	×	×	×
Controls	×	×	×	×
Control $\bar{y}$	0.016	0.156	0.154	0.220
$\mathbb{R}^2$	0.129	0.224	0.224	0.255
Observations	67 737	43 007	54 136	52 069

Notes: Choice fixed effects: year × choice × ag dummies. Controls: age, gender, parental education level and income, and if parents' education is in field j. Standard errors in parentheses.
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

Many cells are dropped since they only contain 1 person, and others have very little variation.

Table A7. Including j/k fixed effects

	Sibling follows to same:			
	(1)	(2)	(3)	(4)
	Choice	Field	Institution	City
Admitted j	0.006*	0.008	0.060***	0.071***
	(0.003)	(0.007)	(0.007)	(0.008)
Choice f.e.	×	×	×	×
Controls	×	×	×	×
Control $\bar{y}$	0.017	0.161	0.170	0.242
$\mathbb{R}^2$	0.273	0.275	0.321	0.348
Observations	45 215	30 609	37 437	36 360

Notes: Choice fixed effects: year × choice × ag dummies for both j and k. Controls: age, gender, parental education level and income, and if parents' education is in field j. Standard errors in parentheses.
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

Table A8 shows the main results but in a model where the choice fixed effects are interacted with gender. This model should capture the variation in admission probability induced by gender-weighted lotteries. Since the model only includes cells where there is at least one randomly admitted and one deferred applicant, many observations are dropped. But still, effects are similar in size and magnitude. However, in Table A9 we run the placebo test with this model and the previously observed correlations persist.

Our last robustness check is in Table A10. Here we instead look at the likelihood that the sibling starts studying the less preferred option k. The dependent variables are now equal to 1 when the sibling's highest ranked choice corresponds to k rather than j. Because z still describes the result of the lottery over j, negative effects indicate following to k.

Sample sizes are smaller because applicants do not always have counterfactual alternatives, for example if they only apply to one choice. Moreover, since k is a less preferred option it is likely that many of the applicants that fail the lottery end up in j anyways. We should thus expect these effects to be weaker.

The two last columns do indeed show the expected effect, on a level almost on par with the main results. That we don't find any effect in the first columns is not surprising, since it was small also in the main specification. Most surprising is the positive result in column (2). While noisy, it fits neatly with the competition hypothesis, since it seems that applicants exhibit anti-following

Table A8. Interacting choice f.e. with gender

	Sibling follows to same:			
	(1)	(2)	(3)	(4)
	Choice	Field	Institution	City
Admitted j	0.009^{***}	0.005	0.047^{***}	0.054^{***}
	(0.002)	(0.006)	(0.005)	(0.006)
Choice f.e.	×	×	×	×
Controls	×	×	×	×
Control $\bar{y}$	0.017	0.160	0.167	0.237
$\mathbb{R}^2$	0.158	0.250	0.258	0.286
Observations	57 198	34 659	44 449	42 587

Notes: Choice fixed effects: year × choice × ag dummies. Controls: age, gender, parental education level and income, and if parents' education is in field j. Standard errors in parentheses.

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

Table A9. Placebo w. gender-interacted f.e.

	Sibling follows to same:			
	(1)	(2)	(3)	(4)
	Choice	Field	Institution	City
Admitted j	0.005^{**}	-0.004	0.008	0.013^{*}
	(0.002)	(0.006)	(0.005)	(0.006)
Choice f.e.	×	×	×	×
Controls	×	×	×	×
Control $\bar{y}$	0.018	0.161	0.161	0.230
$\mathbb{R}^2$	0.201	0.274	0.306	0.334
Observations	50 066	29 024	37 582	35 208

Notes: Choice fixed effects: year × choice × ag dummies. Controls: age, gender, parental education level and income, and if parents' education is in field j. Standard errors in parentheses.Choice f.e. are interacted with a dummy for if the applicant is female.

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

with regards to k, and are more prone to going into the field when their sibling does not.

Table A10. Probability that sibling applies to k

	Sibling follows to same:			
	(1)	(2)	(3)	(4)
	Choice	Field	Institution	City
Admitted j	0.000	0.025^{*}	-0.021^{***}	-0.038^{***}
	(0.002)	(0.010)	(0.006)	(0.008)
Choice f.e.	×	×	×	×
Controls	×	×	×	×
Control $\bar{y}$	0.012	0.119	0.122	0.156
$\mathbb{R}^2$	0.156	0.280	0.242	0.265
Observations	31 979	9658	19 223	17 352

Notes: Choice fixed effects: year × choice × ag dummies. Controls: age, gender, parental education level and income, and if parents' education is in field j. Standard errors in parentheses.
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

References

Altmejd, A. (2017). *Pre Analysis Plan: Effects of Education Choice.* URL: https://osf.io/rj6t7/.

Altonji, J. G., P. Arcidiacono, and A. Maurel (2016). "The Analysis of Field Choice in College and Graduate School: Determinants and Wage Effects." In: *Handbook of the Economics of Education.* Ed. by E. A. Hanushek, S. Machin, and L. Woessmann. Vol. 5. North Holland, pp. 305–396.

Altonji, J. G., E. Blom, and C. Meghir (2012). "Heterogeneity in Human Capital Investments: High School Curriculum, College Major, and Careers." In: *Annual Review of Economics* 4.1, pp. 185–223.

Anelli, M. and G. Peri (2015). "Gender of Siblings and Choice of College Major." In: *CESifo Economic Studies* 61.1, pp. 53–71.

Arcidiacono, P. (2004). "Ability Sorting and the Returns to College Major." In: *Journal of Econometrics* 121.1-2, pp. 343–375.

Beffy, M., D. Fougère, and A. Maurel (2011). "Choosing the Field of Study in Postsecondary Education: Do Expected Earnings Matter?" In: *The Review of Economics and Statistics* 94.1, pp. 334–347.

Berger, M. C. (1988). "Predicted Future Earnings and Choice of College Major." In: *ILR Review* 41.3, pp. 418–429.

Bettinger, E. P., B. T. Long, P. Oreopoulos, and L. Sanbonmatsu (2012). "The Role of Application Assistance and Information in College Decisions: Results from the H&R Block Fafsa Experiment." In: *The Quarterly Journal of Economics* 127.3, pp. 1205–1242.

Betts, J. R. (1996). "What Do Students Know about Wages? Evidence from a Survey of Undergraduates." In: *The Journal of Human Resources* 31.1, pp. 27–56.

Black, S. E. and P. J. Devereux (2011). "Recent Developments in Intergenerational Mobility." In: *Handbook of Labor Economics.* Ed. by D. Card and O. Ashenfelter. Vol. 4, pp. 1487–1541.

Booth, A. and P. Nolen (2012). "Choosing to Compete: How Different Are Girls and Boys?" In: *Journal of Economic Behavior & Organization* 81.2, pp. 542–555.

Bowen, W. G., M. M. Chingos, and M. S. McPherson (2009). *Crossing the Finish Line: Completing College at America's Public Universities.* Princeton University Press.

Brenøe, A. A. (2018). "Sibling Gender Composition and Participation in STEM Education." University of Copenhagen.

Buckles, K., A. Hagemann, O. Malamud, M. Morrill, and A. Wozniak (2016). "The Effect of College Education on Mortality." In: *Journal of Health Economics* 50, pp. 99–114.

Butner, B. et al. (2001). “The College Choice Process of African American and Hispanic Women: Implications for College Transitions.” In: *Journal of College Orientation and Transition* 9.1, pp. 24–32.

Ceja, M. (2006). “Understanding the Role of Parents and Siblings as Information Sources in the College Choice Process of Chicana Students.” In: *Journal of College Student Development* 47.1, pp. 87–104.

Dahl, G. B., K. V. Løken, and M. Mogstad (2014). “Peer Effects in Program Participation.” In: *American Economic Review* 104.7, pp. 2049–2074.

Dustan, A. (2018). “Family Networks and School Choice.”

Eika, L., M. Mogstad, and B. Zafar (2017). *Educational Assortative Mating and Household Income Inequality.* Working Paper 682. Federal Reserv Bank of New York.

Elías McAllister, D. (2012). “Mexican American First-Generation Students’ Perceptions of Siblings and Additional Factors Influencing Their College Choice Process.” PhD Thesis. University of Maryland.

Fagereng, A., M. Mogstad, and M. Rønning (2018). *Why Do Wealthy Parents Have Wealthy Children?* CESifo Working Paper Series 6955.

French, R. and P. Oreopoulos (2017). “Behavioral Barriers Transitioning to College.” In: *Labour Economics*.

Fricke, H., J. Grogger, and A. Steinmayr (2016). “Exposure to Academic Fields and College Major Choice.”

Gelman, A. and E. Loken (2013). “The Garden of Forking Paths: Why Multiple Comparisons Can Be a Problem, Even When There Is No ‘Fishing Expedition’ or ‘p-Hacking’ and the Research Hypothesis Was Posited Ahead of Time.”

Goodman, J., M. Hurwitz, J. Smith, and J. Fox (2015). “The Relationship between Siblings’ College Choices: Evidence from One Million SAT-Taking Families.” In: *Economics of Education Review* 48, pp. 75–85.

Hastings, J., C. A. Neilson, A. Ramirez, and S. D. Zimmerman (2016). “(Un)Informed College and Major Choice: Evidence from Linked Survey and Administrative Data.” In: *Economics of Education Review*. Access to Higher Education 51, pp. 136–151.

Hastings, J., C. A. Neilson, and S. D. Zimmerman (2013). *Are Some Degrees Worth More than Others? Evidence from College Admission Cutoffs in Chile*. Working Paper 19241. National Bureau of Economic Research.

— (2015). *The Effects of Earnings Disclosure on College Enrollment Decisions*. Working Paper 21300. National Bureau of Economic Research.

Högskoleförordningen (1993). SFS 1993:100, until SFS 2017:1326.

Hoxby, C. M. and C. Avery (2012). *The Missing One-Offs: The Hidden Supply of High-Achieving, Low Income Students*. Working Paper 18586. National Bureau of Economic Research.

Jacob, B., B. McCall, and K. Stange (2017). “College as Country Club: Do Colleges Cater to Students’ Preferences for Consumption?” In: *Journal of Labor Economics* 36.2, pp. 309–348.

Kaczynski, K. M. (2011). "Exploring the Influence of Siblings and Their Relationships on the College Choice Process." PhD Thesis. The George Washington University.

Kirkebøen, L. J., E. Leuven, and M. Mogstad (2016). "Field of Study, Earnings, and Self-Selection." In: *The Quarterly Journal of Economics* 131.3, pp. 1057–1111.

Kleinberg, J., H. Lakkaraju, J. Leskovec, J. Ludwig, and S. Mullainathan (2017). "Human Decisions and Machine Predictions." In: *The Quarterly Journal of Economics*.

Konrad, K. A., H. Künemund, K. E. Lommerud, and J. R. Robledo (2002). "Geography of the Family." In: *American Economic Review* 92.4, pp. 981–998.

Kosse, F., T. Deckers, H. Schildberg-Hörisch, and A. Falk (2016). *The Formation of Prosociality: Causal Evidence on the Role of Social Environment*. IZA Discussion Paper Series 9861. IZA.

Lövtrup, M. (2008). "Könsbaserat urval bröt kvinnotrenden." In: *Läkartidningen* 105.51-52, p. 3716.

Manski, C. F. (1993). "Identification of Endogenous Social Effects: The Reflection Problem." In: *Review of Economic Studies* 60.3, p. 531.

Moffitt, R. (2001). "Policy Interventions, Low-Level Equilibria, and Social Interactions." In: *Social Dynamics*. Ed. by H. P. Young and S. Durlauf. Cambridge, MA: MIT Press, pp. 45–82.

Olken, B. A. (2015). "Promises and Perils of Pre-Analysis Plans." In: *The Journal of Economic Perspectives* 29.3, pp. 61–80.

Oreopoulos, P. and K. G. Salvanes (2011). "Priceless: The Nonpecuniary Benefits of Schooling." In: *Journal of Economic Perspectives* 25.1, pp. 159–184.

Pekkala Kerr, S., T. Pekkarinen, M. Sarvimäki, and R. Uusitalo (2015). *Post-secondary education and information on labor market prospects: A randomized field experiment*. IZA Discussion Paper Series 9372. IZA.

Rubin, D. B. (2007). "The Design versus the Analysis of Observational Studies for Causal Effects: Parallels with the Design of Randomized Trials." In: *Statistics in Medicine* 26.1, pp. 20–36.

Schrøter Joensen, J. and H. Skyt Nielsen (2017). *Spillovers in Educational Choice*. SSRN Scholarly Paper ID 2548702. Rochester, NY: Social Science Research Network.

Scott-Clayton, J. (2012). *Information Constraints and Financial Aid Policy*. Working Paper 17811. National Bureau of Economic Research.

Stinebrickner, R. and T. Stinebrickner (2014). "A Major in Science? Initial Beliefs and Final Outcomes for College Major and Dropout." In: *The Review of Economic Studies* 81.1, pp. 426–472.

Svensson, L.-G. (1999). "Strategy-Proof Allocation of Indivisible Goods." In: *Social Choice and Welfare* 16.4, pp. 557–567.

Webber, D. A. (2014). “The Lifetime Earnings Premia of Different Majors: Correcting for Selection Based on Cognitive, Noncognitive, and Unobserved Factors.” In: *Labour Economics* 28, pp. 14–23.