TI 2023-056/V
Tinbergen Institute Discussion Paper

Spillovers in Fields of Study: Siblings, Cousins, and Neighbors

Stanislav Avdeev[1,3]
Nadine Ketel[1,3]
Hessel Oosterbeek[1,3]
Bas van der Klaauw[1,3]

1 University of Amsterdam
2 Vrije Universiteit Amsterdam
3 Tinbergen Institute

Spillovers in Fields of Study: Siblings, Cousins, and Neighbors*

Stanislav Avdeev, Nadine Ketel,
Hessel Oosterbeek, Bas van der Klaauw

Abstract

We use admission lotteries for higher education studies in the Netherlands to investigate whether someone's field of study influences the study choices of their younger peers. We find that younger siblings and cousins are strongly affected. Also younger neighbors are affected but to a smaller extent. These findings indicate that a substantial part of the correlations in study choices between family members can be attributed to spillover effects and are not due to shared environments. Our findings contrast with those of recent studies based on admission thresholds, which find no sibling spillovers on field of study (major) choices. Because we also find spillovers from lottery participants at the lower end of the ability distribution, the contrasting findings cannot be attributed to the different research designs (leveraging admission lotteries versus admission thresholds). We believe that the different findings are due to the small differences in quality between universities in the Netherlands, making differences in the prestige of fields of study more prominent.

Keywords: Major choice; Higher education; Peer effects; Admission lotteries.

JEL Codes: I23, I24, J10

*This version: September 2023. Avdeev: University of Amsterdam (s.avdeev@uva.nl); Ketel: VU University Amsterdam (n.ketel@vu.nl); Oosterbeek: University of Amsterdam (h.oosterbeek@uva.nl); Van der Klaauw: VU University Amsterdam (b.vander.klaauw@vu.nl).
We gratefully acknowledge valuable comments from conference and seminar participants in Amsterdam, Belgrade, Catanzaro, Leuven and Oslo. Ketel acknowledges financial support from a Veni-grant from the Dutch Science Foundation (NWO). Oosterbeek received support from the Research Council of Norway Toppforsk grant no. 275906. The non-public micro data used in this paper are available via remote access to the Microdata services of Statistics Netherlands (project agreement 7930).

1 Introduction

Choosing a field of study is among the most consequential choices that young people make. It is important for their prospects in the labor market (Kirkeboen et al., 2016; Ketel et al., 2016, 2019; Bleemer and Mehta, 2022), in the marriage market (Artmann et al., 2021; Kirkeboen et al., 2022), and for their health (Leuven et al., 2013). Yet, little is known about the factors that determine field of study (or major) choices. This is exemplified by the conclusion in Wiswall and Zafar (2015) who state that "even with our rich data on beliefs across a variety of pecuniary and non-pecuniary aspects of majors, major-choices in our data are still largely the result of heterogeneity in major specific and unobserved "tastes"."[1] Information about the factors that determine field of study choices is an important input for policies that aim to influence these choices.

Positive correlations between study choices of family members and of neighbors suggest that network spillovers play a role for study choices (Raaum et al., 2003; Hällsten, 2014; Goodman et al., 2015; van der Vleuten et al., 2020). Separating such spillovers from the mere correlation due to a shared environment is, however, challenging.

A handful of recent studies has made progress in separating spillover effects from the effects of a shared environment by analyzing higher-education choices of the younger siblings and younger neighbors of applicants who were close to an admission threshold (Altmejd et al., 2021; Aguirrea and Matta, 2021) or a loan eligibility threshold (Barrios-Fernández, 2022) in regression discontinuity designs.[2] Specifically, Altmejd et al. (2021) study sibling spillovers in college and

[1] Wiswall and Zafar (2015) conduct a survey experiment in which they provide information about major-specific characteristics to study the determinants of college major choice.

[2] There is also work on spillovers in study choices in secondary education. Older siblings influence their younger siblings' preferred high school in Mexico (Dustan, 2018), take up of advanced math and science courses in Denmark (Joensen and Nielsen, 2018), secondary education specialization in the Netherlands (van der Vleuten et al., 2020), take up of high school Advanced Placement exam in the US (Gurantz et al., 2020), and high school major in Sweden (Dahl et al., 2021). Some papers look at sibling spillovers on achievement (Oettinger, 2000; Qureshi, 2018a,b; Nicoletti and Rabe, 2019; de Gendre, 2021; Karbownik and Özek, 2021; Zang et al., 2023; Figlio et al., 2023; Goldstein, 2023). Two Swedish studies find strong parent spillovers on children's high school major (Dahl et al., 2021) and university field of study (Altmejd, 2023). There is some evidence of positive neighbor spillovers on secondary school enrollment in Mexico (Bobonis and Finan, 2009; Lalive and Cattaneo, 2009) and Chile (Matta and Orellana, 2022). Neighbors also influence early and secondary education performance in

major choices in Chile, Croatia, Sweden, and the US. The authors use admission thresholds to compare younger siblings' study choices whose older siblings were just above or below the thresholds. They find that younger siblings follow their older siblings to the same college (institution), but not to the same major (field of study). Similarly, Aguirrea and Matta (2021) exploit admission thresholds generated by Chile's centralized higher education system and find strong sibling spillover effects on university enrollment but not on field of study choices. Barrios-Fernández (2022) exploits variation due to student loans eligibility in Chile. He finds that neighbors are more likely to attend and complete a university if their closest neighbor becomes eligible for a student loan.[3]

The current paper expands the evidence on spillovers in field of study choices by leveraging variation caused by admission lotteries to higher education studies in the Netherlands. During the period from 1987 to 1999, admission to oversubscribed fields of study was determined by the results from centralized admission lotteries. Fields of study in the health sector that ran admission lotteries are medicine, veterinary medicine, dentistry, occupational therapy, and biology. Other fields of study that ran admission lotteries are business, international business, and tourism. Due to the admission lotteries, some of the applicants to these fields of study were admitted while other, equally qualified, applicants were rejected. A large share of the rejected applicants ended up in another field of study. The rich administrative data from Statistics Netherlands allow us to link information from lottery participants to the study choices of their younger siblings as well as to those of their younger cousins and their younger neighbors. We can therefore analyze spillovers of fields of study on three different types of younger peers in one framework.[4]

Our main finding is that there are strong spillovers of the fields of study of the lottery participants on the fields of study choices of their younger peers. Relative to a baseline (control complier mean) of 2.9%, the younger siblings of lottery

France (Goux and Maurin, 2007) and the US (List et al., 2020). Goulas et al. (2022) find that a higher share of female neighbors increases enrollment in an academic university and female enrollment in a STEM field.

[3]There are some papers studying correlations of cousins' GPA and years of education in the US (Jæger, 2012) and in Sweden (Hällsten, 2014; Adermon et al., 2021; Collado et al., 2022; Hällsten and Kolk, 2022), but no papers studying cousin spillovers on the field of study choices.

[4]Barrios-Fernández (2022) estimates spillovers of siblings and close neighbors (but not cousins) on university enrollment (but not on field of study) in the context of Chile.

participants who enrolled in the lottery field of study because they won the first lottery are 5.7 percentage points more likely to enroll in that same field of study. For younger cousins this effect is 2.2 percentage points (relative to a baseline of 1.1%), and for younger neighbors it is 0.3 percentage points (relative to a baseline of 1%).

Sibling spillovers are larger when the lottery participant and the younger sibling are of the same sex, suggesting a role model effect. There is no such interaction effect for cousins or neighbors. Neighbor spillovers are larger when the age difference between the lottery participant and the younger neighbor is smaller. Sibling and cousin spillovers are larger in high-income families suggesting a role of parental resources.

Our finding of substantial sibling spillovers on field of study choices deviates from the results of Altmejd et al. (2021) and Aguirrea and Matta (2021). To inquire whether this is because we leverage admission lotteries instead of admission thresholds, we estimate spillovers of applicants from the lowest ability (based on secondary school GPA) group. Spillovers from these applicants are not different from those of other applicants, indicating that the contrasting findings cannot be attributed to the use of different research designs. A possible reason for the different results is that in the Dutch setting all institutions (universities and colleges) are publicly and equally funded, resulting in small, if any, quality differences between them. This makes differences in the prestige (perceived difficulty) of fields of study more important, for example as a signal to future employers. This situation is quite different from those in Chile, Croatia, Sweden and the US where some institutions have much stronger reputations than others.

The paper continues as follows. Section 2 describes the Dutch higher education system and its use of admission lotteries. Section 3 describes the data. Section 4 discusses details of the identification strategy. Section 5 presents and discusses the results. Section 6 summarizes and concludes.

2 Institutional background

Students in the Netherlands who completed the six-years university track or the five-years college track in secondary school can enroll in any Dutch university or college (also referred to as university of applied sciences). Unlike the situation in the US, students choose their field of study upon application. Most fields of study accept all applicants, while some fields of study put a cap (quota) on the number of admitted students. Until 1999, fields of study with a quota held nationwide centralized lotteries to admit students. After 1999, the centralized admission lotteries were gradually replaced by decentralized admission policies. The fields of study that held at least one admission lottery between 1988 (the first year for which we can use data) and 1999 are: medicine, veterinary medicine, dentistry, occupational therapy, biology, business, international business, and tourism. The admission lotteries for these fields of study pertain to all universities or colleges that offer them.[5]

Lotteries were introduced as a response to increasing numbers of applicants, which exceeded the capacities of the fields of study. The Ministry of Education determined these capacities. To strike a balance between merit and equality of opportunity, the lotteries gave students with a higher GPA on the nationwide final exams from secondary school higher chances to be admitted. GPA is expressed on a scale from 1 to 10, where 6.0 or above indicates a pass. Table 1 shows which GPA intervals are assigned to the different lottery categories, labeled A to F.[6] The table also shows the admission probabilities and shares of applicants for each category averaged across different fields of study with a lottery and lottery years. Only 5% of the applicants are in the top two categories which have admission probabilities above 70%. All other applicants are exposed to much higher chances to lose the lottery of the field of study of their choice.

An important feature of the admission lotteries is that lottery losers are allowed to participate in subsequent lotteries in later years. We address this feature by using the results of the first lottery in which someone participated as an instrumental variable for enrollment in the field of study with the lottery.[7] Leuven et al. (2013) and Ketel et al. (2016) provide more detailed descriptions of the Dutch admission lotteries and educational system.

[5] There have also been admission lotteries for some fields of study at specific universities or colleges. Because the losers of these institution×field of study-lotteries could enroll in the same field of study elsewhere, we do not use these lotteries in our analyses.

[6] Category "Other" refers to students who did not attend secondary school in the Netherlands and therefore did not participate in the final school exams, such as foreign students.

[7] It is not allowed to participate in multiple admission lotteries in the same year.

Table 1: Lottery categories

Category	GPA	Share	P(win)
A	GPA $\geq$ 8.5	0.01	0.87
B	$8.0 \leq$ GPA < 8.5	0.04	0.77
C	$7.5 \leq$ GPA < 8.0	0.08	0.65
D	$7.0 \leq$ GPA < 7.5	0.20	0.56
E	$6.5 \leq$ GPA < 7.0	0.24	0.48
F	GPA < 6.5	0.32	0.43
Other	–	0.11	0.56

Notes: GPA is grade point average on the nationwide final exams in secondary school. Share is the share of applicants in each category. P(win) indicates the probability of being admitted in each category. Share and P(win) are averaged across different fields of study with a lottery and lottery years. The category "Other" refers to students who did not participate in the secondary school exams, such as foreign students. This category is excluded from the analysis.

3 Data

3.1 Data source and sample

We use administrative data from different registers available at Statistics Netherlands, which can be linked at the individual level, at the parent-child level, and at the neighborhood level. The person and household registers contain information on different demographic and education variables, i.e., age, sex, immigration background, neighborhood, university or college enrollment, and field of study. The register on admission lotteries contains information on all lottery participants, their lottery category, the year of participation, and the lottery results. We use first-time lottery participants between 1988 and 1999. We exclude applicants older than 21 when applying for the first time. Applicants from top category A are dropped because almost all of them are admitted. Finally, we exclude first-generation immigrants because it is often impossible to trace their cousins.

The units of the analysis in our study are sibling pairs, cousin pairs, and neighbor pairs. Siblings are individuals who have one or two parents in common. Cousins are individuals who have a common grandparent. Neighbors are individuals who live in the same neighborhood. We construct three samples.

To construct the samples of siblings and of cousins, we link lottery participants to their younger siblings and cousins using the registry of their (grand-)parents. We restrict the sample to lottery participants who have at least one sibling or cousin who is at least one year younger. If there are several older siblings or cousins participating in a lottery, we take the oldest older sibling or cousin to determine the treatment status of the younger ones. Results are almost identical when we use the youngest older sibling or cousin, or when we exclude families with more than one older sibling or cousin participating in a lottery. Our analyses are based on more than 33 thousand sibling pairs and on almost 80 thousand cousin pairs.

For the sample of neighbors, we link lottery participants to their neighbors using neighborhood codes ("buurtcodes"). This information is available from 1995 onward. For the analyses of neighbor spillovers, we therefore only use lottery participants from 1995 onward. The average number of inhabitants per neighborhood is 1400. We exclude siblings and cousins from the sample of neighbors. To keep the analysis for neighbors similar to those for siblings and cousins, we restrict the sample to neighborhoods where only one person participated in a lottery. This leaves us with 48% of the lottery participants and 71% of the neighborhoods.[8] We consider spillovers on neighbors who are between one and two years younger than the lottery participant.[9] On average, there are 28 of such younger neighbors per neighborhood. Our analyses are based on more than 317 thousand pairs of neighbors.

3.2 Descriptive statistics

Panel A of Table 2 reports mean values of predetermined characteristics of lottery participants by the outcome of the first lottery. Around 55% of the lottery participants are female, 3% have a non–western background, and the average age at the first lottery is 18.8. Lottery participants have on average 2.5 younger siblings, 5.2 younger cousins and 27.6 younger neighbors. The average parental

[8]In Subsection 5.4 we report results from an analysis where we also include neighborhoods with multiple lottery participants.

[9]The minimum age difference of one year is chosen to avoid reflection issues. The maximum age difference of two years is chosen to increase the likelihood that neighbors interact with each other.

income of lottery participants is 73 thousand euro.

Panels B, C, and D of Table 2 report mean values of the characteristics of the younger peers of the lottery participants. Almost 50% of the younger peers are female and around 5% have a non-western background. The average age difference between older and younger siblings is 4.6 years, between older and younger cousins 9.2 years, and between older and younger neighbors 1.5 years.[10] 50% of all pairs are same-sex pairs.

The final column of Table 2 reports p-values obtained from regressing personal characteristics on an indicator for the older peer winning the first lottery, fixed effects for lottery field of study, lottery category and year of the first lottery participation, and interactions of these fixed effects. The p-values indicate that the samples of lottery winners and lottery losers themselves as well as the samples of their younger siblings, cousins and neighbors, are balanced.

The lottery participants are divided over the different fields of study as follows: medicine 57.9%, veterinary medicine 9.4%, dentistry 4.8%, occupational therapy 3.8%, biology 1%, international business 17.7%, tourism 3.1%, and business 2.3%. While these percentages indicate that the larger share of applicants participated in the lottery for medicine or in a lottery in a health-related field, we find similar results when we exclude medicine or all health-related fields of study from the analysis.

4 Empirical approach

To estimate the effect of an older peer's enrollment in the lottery field of study on the younger peer's study choice, we use the following model:

$$Y_{ip} = \delta D_{ip} + X_i' \beta + \mu_p + \varepsilon_{ip}, \quad (1)$$

where Y_{ip} indicates whether the younger peer in pair i is enrolled in the field of study p for which the older peer participated in a lottery. D_{ip} indicates whether the older peer in pair i is enrolled in the lottery field of study p. X_i is a vector of control variables including dummies for sex and non-western origin, and age

[10]This last number is due to the sample restriction of younger neighbors being between one and two years younger.

Table 2: Balancing table

	Lottery winners	Lottery losers	p-value
Panel A: Lottery participants			
Female	0.55	0.56	0.05
Non-western immigrant	0.03	0.03	0.58
Age at first application	18.84	18.88	0.36
Number of younger siblings	2.48	2.50	0.88
Number of younger cousins	5.24	5.24	0.54
Number of younger neighbors	27.79	27.49	0.49
Parental income	70,133	74,853	0.27
N	18,588	20,058	
Panel B: Younger siblings			
Female	0.49	0.48	0.25
Non-western immigrant	0.04	0.05	0.98
Age difference in years	4.63	4.58	0.64
Same sex pairs	0.49	0.49	0.81
Parental income	81,018	81,805	0.05
N	15,677	17,417	
Panel C: Younger cousins			
Female	0.49	0.49	0.37
Non-western immigrant	0.02	0.02	0.11
Age difference in years	9.22	9.23	0.48
Same sex pairs	0.50	0.50	0.92
Parental income	69,905	69,876	0.59
N	38,012	41,692	
Panel D: Younger neighbors			
Female	0.50	0.50	0.06
Non-western immigrant	0.08	0.08	0.05
Age difference in years	1.45	1.45	0.17
Same sex pairs	0.50	0.50	0.13
Parental income	55,647	56,734	0.20
N	131,260	185,957	

Notes: Since the lottery is weighted, the observed differences between lottery winners and losers in this table cannot be interpreted as causal. p-values obtained from regressing personal characteristics on an indicator for the older peer winning the first lottery, including fixed effects for lottery field of study, lottery category and year of the first lottery participation, and interactions of these fixed effects. All regressions use the normalized inverse of the propensity score as a weight.

of the older peer at the first lottery. These variables are included to increase precision. μ_p are fixed effects for lottery field of study, lottery category and year of the first lottery participation, and their interactions.

Because not all lottery participants comply with the result of the first lottery in which they participate, estimation of equation (1) with OLS, may result in a biased estimate of δ.[11] We therefore instrument D_{ip} with the result of the first lottery in which someone participated (Z_{ip}). We estimate a first-stage equation of the form:

$$D_{ip} = \alpha Z_{ip} + X_i' \gamma + \lambda_p + \nu_{ip}, \tag{2}$$

where α indicates the compliance rate, i.e., the difference in enrollment rates between winners and losers of the first admission lottery in which they participated. Observations are weighted by the normalized inverse of the propensity score to win the lottery, such that estimates of δ can be interpreted as (local) average treatment effects (cf. Słoczyński et al., 2022). We cluster standard errors at the level of the older peer as the treatment is assigned at this level (cf. Abadie et al., 2023).

5 Results

We present the results in four parts. We start with the first-stage results. Next we present the main results of spillovers on younger peers. The third subsection presents our findings regarding heterogeneity of spillovers, which are informative about (the absence of) certain mechanisms. The final subsection discusses findings from various robustness analyses.

5.1 First-stage results

Table 3 shows the estimates of the effect of winning the first lottery on the probability to enroll in the lottery field of study. The table has two panels. Panel A reports results based on a sample of older peers where each older peer appears once, provided that they have at least one younger sibling, cousin or neighbor.

[11] Compliance is imperfect because not all winners enroll in the lottery field of study and because lottery losers often reapply and enroll in subsequent years in the lottery field of study.

Panel B reports results based on a sample of older peers where each older peer appears as often as they have younger siblings, cousins or neighbors. The results in Panel A are informative about the strength of the first-stage relationship. The results in Panel B are the relevant input for the IV regressions. It is reassuring that the results in the two panels are very similar.

The control means in Table 3 indicate that 42-47% of the applicants who lose their first admission lottery enroll in the lottery field of study. This occurs after winning an admission lottery in a later year. Winning the first lottery increases the probability to be enrolled in the lottery field of study by around 50 percentage points.[12] Hence, winning the first lottery has a strong effect on the probability to be enrolled in the lottery field of study.

Table 3: First-stage estimates of winning the first lottery on enrollment in the lottery field of study

	Lottery participant enrolls in the lottery field of study		
	Sibling	**Cousin**	**Neighbor**
Panel A: Unique older peers			
Lottery participant wins	0.502***	0.506***	0.469***
	(0.006)	(0.006)	(0.008)
Control mean	0.434	0.426	0.471
F-statistic	9,853.13	8,308.46	4,089.84
N	22,201	18,808	10,953
Panel B: Duplicated older peers			
Lottery participant wins	0.495***	0.507***	0.476***
	(0.006)	(0.007)	(0.010)
Control mean	0.439	0.422	0.460
F-statistic	14,462.46	36,971.54	124,138.84
N	33,094	79,704	317,217

Notes: All regressions include controls for sex, non-western origin, age of the older peer at the first lottery, and fixed effects for lottery field of study, lottery category, year of the first lottery participation, and their interactions. Panel A includes a sample of older peers where each older peer appears once. Panel B includes a sample of older peers where each older peer appears as often as they have younger peers. Standard errors in Panel B are clustered at the older peer level (the lottery participant). *p<0.1; **p<0.05; ***p<0.01

[12] The first-stage F-statistic is well above 104.7 in all models. These values ensure that reported IV standard errors are valid for the 0.05 significance level (Lee et al., 2022).

5.2 Main results

Panel A of Table 4 presents the main results of this paper. The first column presents the IV estimate of sibling spillovers, the second of cousin spillovers and the third of neighbor spillovers.

Turning first to sibling spillovers, we find that if an older sibling enrolls in the lottery field of study because they won the first lottery, a younger sibling's probability to enroll in the same field of study increases by 5.7 percentage points. This is a substantial effect in comparison to the 2.9% enrollment rate of the younger siblings of losing compliers. This base rate of 2.9% is only slightly above the mean enrollment rate in the population of younger siblings of higher education students from the same cohorts.

Column (2) shows the results for spillovers on younger cousins. We find an increase of 2.2 percentage points on the probability that the younger cousin enrolls in the lottery field of study if the older cousin enrolled in it. This should be compared to a base rate (control complier mean) of 1.1%.

Column (3) presents the results for neighbor spillovers. There is a small but statistically significant effect of older neighbors' enrollment in the lottery field of study on younger neighbors' enrollment in the same field of study: 0.3 percentage points compared to a base rate of 1%.

The estimated spillover effects are larger for siblings than for cousins, and larger for cousins than for neighbors. When we multiply the respective estimates with the numbers of people exposed to them, the total spillover on siblings amounts to 0.141 (=0.057×2.48), the total spillover on cousins equals 0.115 (=0.022× 5.24), and the total spillover on neighbors is 0.083 (=0.003×27.79). Hence, the multiplier effect through siblings is about twice as large as the multiplier effect through neighbors. The multiplier effect through cousins is in between these two.

Our findings of positive sibling spillovers on field of study choices deviate from the results from the recent studies of Altmejd et al. (2021) and Aguirrea and Matta (2021). A first possible reason for the different results is that we use admission lotteries for identification whereas the other studies use admission thresholds. This means that we estimate spillovers of older siblings from the entire ability distribution whereas the other studies estimate spillovers of older

Table 4: Spillover effect of older peer's lottery result/enrollment in the lottery field of study on younger peer's study choice

	Younger peer enrolls in the lottery field of study		
	Sibling	**Cousin**	**Neighbor**
Panel A: Main results			
Older peer wins the first lottery	0.028***	0.011***	0.001***
	(0.003)	(0.001)	(0.001)
Older peer enrolls	0.057***	0.022***	0.003***
	(0.005)	(0.002)	(0.001)
Control complier mean	0.029	0.011	0.010
N	33,094	79,704	317,217
Panel B: Heterogeneity by lottery category			
Older peer enrolls	0.061***	0.026***	0.003**
	(0.007)	(0.003)	(0.001)
Older peer enrolls	-0.008	-0.009**	-0.001
x $\mathbb{1}$ [Older peer's GPA < 6.5]	(0.010)	(0.005)	(0.002)
N	33,094	79,704	317,217
Panel C: Heterogeneity by sex composition			
Older peer enrolls	0.050***	0.021***	0.003**
	(0.007)	(0.003)	(0.001)
Older peer enrolls	0.020**	0.001	-0.001
x $\mathbb{1}$ [Same sex = 1]	(0.010)	(0.004)	(0.002)
N	33,094	79,704	317,217
Panel D: Heterogeneity by age gap			
Older peer enrolls	0.059***	0.024***	0.003***
	(0.006)	(0.004)	(0.001)
Older peer enrolls	-0.007	-0.003	-0.003***
x $\mathbb{1}$ [Age gap > 5 (2)]	(0.011)	(0.004)	(0.001)
N	33,094	79,704	1,472,910
Panel E: Heterogeneity by parental income			
Older peer enrolls	0.041***	0.017***	0.001
	(0.005)	(0.002)	(0.001)
Older peer enrolls	0.035***	0.012**	0.003
x $\mathbb{1}$ [Parental income > median]	(0.011)	(0.005)	(0.002)
N	32,995	79,484	316,141

Notes: All regressions include controls for sex, non-western origin, age of the older peer at the first lottery, and fixed effects for lottery field of study, lottery category, year of the first lottery participation, and their interactions. All regressions use the normalized inverse of the propensity score as a weight. In Panel D, we include interactions with an indicator equal to 1 if the age difference is larger than 5 for siblings and cousins, and larger than 2 for neighbors. In Column (3) of Panel D, we include neighbors with at most a 10 years age gap. In Panel E, we include interactions with an indicator equal 1 if the parental income of older peers is above the sample median. Standard errors are clustered at the older peer level. *p<0.1; **p<0.05; ***p<0.01

siblings from the lower end of the ability distribution. It is conceivable that the latter have less favorable experiences in their studies than the average student and therefore generate no or smaller spillovers.

We can test this explanation by estimating whether the spillovers from lottery applicants with the lowest GPA on the secondary school exams (GPA < 6.5) differ from those of the other applicants. The results in Panel B of Table 4 indicate that we cannot reject that sibling and neighbor spillovers from low-ability applicants are the same as those from other applicants. Cousin spillovers from low-ability applicants are smaller but still substantial. The difference in research designs is therefore an unlikely explanation for the different results.

An alternative explanation are institutional differences. Unlike the situation in Chile, Croatia, Sweden, and the US (the countries included in the other studies), there are only small differences in the quality and reputation of Dutch universities. This is illustrated by data from the Times Higher Education World University Ranking of 2023. The standard deviation of the ranks of Dutch universities in this ranking equals 81 (with a mean rank of 128). For universities in Chile, Croatia, Sweden and the US the standard deviations of their ranks are 299, 202, 245 and 374, respectively.[13]

Reasons for the small relative differences in quality and reputation between universities in the Netherlands are that all universities and colleges are publicly and equally funded, that there is no selection of applicants on grades, and that salaries of faculty are determined by collective agreements. There is therefore no prestige attached to choosing one particular university instead of another. This is different for the choice of fields of study – which students do immediately when they apply. Some fields of study are known to be very challenging (physics, econometrics) whereas others are less so (business administration, communication science). This translates in different fields of study attracting students from different parts of the ability distribution. In short: in the Dutch setting the choice of a field of study is more salient than the choice of an institution.

[13]The mean rank of universities in the Netherlands is 128, for Chile 1211, for Croatia 1317, for Sweden 330 and for the US 476.

5.3 Heterogeneous spillovers

We now turn to heterogeneity in the spillover effects, which shed light on (the absence of) certain mechanisms underlying spillovers. We first examine whether spillovers are different for same-sex peers than for opposite-sex peers. This may be informative about role models as a driver of peer effects. Next, we investigate whether spillovers vary with the age gap between the peers. Finally, we explore whether spillovers differ between families with above and below median income.

Sex composition. Panel C of Table 4 presents spillovers by the sex composition of pairs. The first column shows that the spillover between same-sex siblings is 2.0 percentage points larger than the 5.0 percentage points spillover between opposite-sex siblings. There are no such differences for cousins and neighbors. The result for siblings lends support for a role model mechanism, because older siblings can serve as role models for their younger siblings (Bandura, 1977; McHale et al., 2012) which is more likely for same-sex siblings than for opposite-sex siblings.

Age gap. Panel D of Table 4 presents results of spillovers that vary with the age gap between peers. For siblings and cousins we interact enrollment in the lottery field of study with a dummy for being more than 5 years apart, and for neighbors for being more than 2 years apart. For siblings and cousins the results reveal no variation of spillovers with the age gap. For neighbors, there is basically no spillover when the age gap exceeds two years. The results for siblings and cousins are at odds with an information mechanism, which predicts that spillovers are stronger when peers are closer because information is more up-to-date (Festinger, 1954; Whiteman et al., 2011). The result for neighbors is consistent with the information channel. It might, however, also be the case that neighbors who are more than two years apart have little interaction.

Parental income. Panel E of Table 4 investigates differences in spillovers between families where parental income is above or below the median. While the sibling spillover is 4.1 percentage points for families with below median income, it is 7.6 percentage points for families with above median income. For cousins spillovers, the respective numbers are 1.7 percentage points and 2.9 percentage points. There are no differential neighbor spillovers by income. A possible mechanism that drives the differential spillovers by income for siblings and cousins is that the older peers from higher income families have made more informed choices than the older peers from lower income families. Alternatively, it may be that the older peers from higher income families enjoy studying more because they have to borrow less or do not need a side job to finance their studies.

5.4 Robustness

Table A1 in the appendix, presents results separately for all lottery fields of study except medicine, and for the three lottery fields of study that are not health-related (business, international business and tourism). The resulting estimates are a bit smaller and less precise than the main results but support to the same conclusions.

Table A2 in the appendix presents estimates of the effect of the older peer winning the first lottery on the probability of the younger peer to enroll in any higher education study. Only in the cousin sample there is some indication that the result of the first lottery has an extensive margin effect. This effect is, however, only significant at the 10%-level and small in comparison to the control complier mean.

Table A3 in the appendix presents estimates of spillovers using different sample definitions for siblings, cousins, and neighbors. In the main analysis, we use the oldest older sibling to determine the treatment status of all younger siblings, and these are the baseline results in Column (1) of Panel A of Table A3. In Column (2), we use the youngest older sibling, and in Column (3) – we use only families with one older sibling participating in a lottery. Overall, the estimates are virtually unchanged. We repeat this for cousins. In the main analysis, we use the oldest older cousin to determine the treatment status of younger ones, and these are baseline results in Column (1) of Panel B of Table A3. In Column (2), we use the youngest older cousin, and in Column (3) – we use only extended families with one older cousin participating in a lottery. The results are almost identical across different samples.

To keep the results for neighbors comparable to those of siblings and cousins, we restricted the neighbor sample to neighborhoods with one lottery participant. To also include neighborhoods with multiple lottery participants in the analysis, we use the following model (Abdulkadiroğlu et al., 2017):

$$Y_{in} = \delta D_{in} + p_{in} + \varepsilon_{in}$$
$$D_{in} = \alpha Z_{in} + p_{in} + \nu_{in}$$

where Y_{in} indicates whether the younger neighbor i is enrolled in the field of study for which at least one older neighbor participated in an admission lottery. D_{in} indicates whether there is at least one older neighbor in neighborhood n who enrolled in that field of study, Z_{ip} indicates whether at least one older neighbor in neighborhood n won their first admission lottery. p_{in} is the assignment propensity score that at least one older neighbor in neighborhood n is a lottery winner. We include this score as a fixed effect. This extension of the model takes account of the fact that neighborhoods differ in the number of lottery participants and therefore in the probability of exposure to a lottery winner. We weight observations with the normalized inverse of the propensity score. Standard errors are clustered at the neighborhood level. Column (1) of Panel C of Table A3 presents the main results using only neighborhoods with one lottery participant, in Column (2) we use neighborhoods with multiple lottery participants with the propensity score strictly between 0 and 1, and in Column (3) - with the propensity score strictly between 0.1 and 0.9. The results are very similar.

6 Conclusion

Leveraging admission lotteries for higher education studies in the Netherlands, we find that the study choices of younger siblings and cousins are strongly affected by the field of study in which their older sibling or cousin enrolls. Also younger neighbors are affected but to a smaller extent. These findings indicate that a substantial part of the correlations in study choices between family members can be attributed to spillover effects and are not due to shared environments.

Our findings diverge with those of recent studies based on admission thresholds, which find no sibling spillovers on field of study choices. These contrasting findings cannot be attributed to the different research designs. We believe that the different findings are probably due to the small differences in quality between universities and colleges in the Netherlands, making differences in the prestige of fields of study more prominent.

Our results suggest that policies aimed at influencing young people's study choices, have multiplier effects, mainly through family networks. Our findings concerning differential effects by sex composition and family income should be taken into consideration when designing policies so that possible negative side effects regarding the diversity of the student body can be avoided.

References

Abadie, A., S. Athey, G. W. Imbens, and J. M. Wooldridge (2023). When Should You Adjust Standard Errors for Clustering? *The Quarterly Journal of Economics 138*(1), 1–35.

Abdulkadiroğlu, A., J. D. Angrist, Y. Narita, and P. A. Pathak (2017). Research Design Meets Market Design: Using Centralized Assignment for Impact Evaluation. *Econometrica 85*(5), 1373–1432.

Adermon, A., M. Lindahl, and M. Palme (2021). Dynastic Human Capital, Inequality, and Intergenerational Mobility. *American Economic Review 111*(5), 1523–1548.

Aguirrea, J. and J. Matta (2021). Walking in Your Footsteps: Sibling Spillovers in Higher Education Choices. *Economics of Education Review 80*, 1–28.

Altmejd, A. (2023). Inheritance of Fields of Study. *Working Paper*, 1–55.

Altmejd, A., A. Barrios-Fernández, M. Drlje, J. Goodman, M. Hurwitz, D. Kovac, C. Mulhern, C. Neilson, and J. Smith (2021). O Brother, Where Start Thou? Sibling Spillovers on College and Major Choice in Four Countries. *The Quarterly Journal of Economics 136*(3), 1831–1886.

Artmann, E., N. Ketel, H. Oosterbeek, and B. van der Klaauw (2021). Field of Study and Partner Choice. *Economics of Education Review 84*(102149), 1–7.

Bandura, A. (1977). *Social Learning Theory*. Prentice Hall: Englewood Cliffs.

Barrios-Fernández, A. (2022). Neighbors' Effects on University Enrollment. *American Economic Journal: Applied Economics 14*(3), 30–60.

Bleemer, Z. and A. Mehta (2022). Will Studying Economics Make You Rich? A Regression Discontinuity Analysis of the Returns to College Major. *American Economic Journal: Applied Economics 14*(2), 1–22.

Bobonis, G. J. and F. Finan (2009). Neighborhood Peer Effects in Secondary School Enrollment Decisions. *The Review of Economics and Statistics 91*(4), 695–716.

Collado, M. D., I. Ortuño-Ortín, and J. Stuhler (2022). Estimating Intergenerational and Assortative Processes in Extended Family Data. *The Review of Economic Studies*, 1–33.

Dahl, G., D.-O. Rooth, and A. Stenberg (2021). Intergenerational and Sibling Peer Effects in High School Majors. *NBER Working Paper 27618*, 1–30.

de Gendre, A. (2021). Class Rank and Sibling Spillover Effects. *Working Paper*, 1–57.

Dustan, A. (2018). Family Networks and School Choice. *Journal of Development Economics 134*, 372–391.

Festinger, L. (1954). A Theory of Social Comparison Processes. *Human Relations 7*(2), 117–140.

Figlio, D. N., K. Karbownik, and U. Özek (2023). Sibling Spillovers May Enhance the Efficacy of Targeted School Policies. *NBER Working Paper 31406*, 1–19.

Goldstein, Y. (2023). Learning from Round Test Scores: University Field of Study Decisions and Career Outcomes. *Working Paper*, 1–36.

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

Goulas, S., R. Megalokonomou, and Y. Zhang (2022). Female Neighbors, Test Scores, and Careers. *IZA Discussion Papers 15733*, 1–66.

Goux, D. and E. Maurin (2007). Close Neighbours Matter: Neighbourhood Effects on Early Performance at School. *The Economic Journal 117*(523), 1193–1215.

Gurantz, O., M. Hurwitz, and J. Smith (2020). Sibling Effects on High School Exam Taking and Performance. *Journal of Economic Behavior and Organization 178*, 534–549.

Hällsten, M. (2014). Inequality across Three and Four Generations in Egalitarian Sweden: 1st and 2nd Cousin Correlations in Socio-Economic Outcomes. *Research in Social Stratification and Mobility 35*, 19–33.

Hällsten, M. and M. Kolk (2022). The Shadow of Peasant Past: Seven Generations of Inequality Persistence in Northern Sweden. *Working Paper*, 1–113.

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

Jæger, M. M. (2012). The Extended Family and Children's Educational Success. *American Sociological Review 77*(6), 903–922.

Karbownik, K. and U. Özek (2021). Setting a Good Example? Examining Sibling Spillovers in Educational Achievement Using a Regression Discontinuity Design. *Journal of Human Resources*, 1–13.

Ketel, N., E. Leuven, H. Oosterbeek, and B. van der Klaauw (2016). The Returns to Medical School: Evidence from Admission Lotteries. *American Economic Journal: Applied Economics 8*(2), 225–254.

Ketel, N., E. Leuven, H. Oosterbeek, and B. Van Der Klaauw (2019). Do Dutch Dentists Extract Monopoly Rents? *Journal of Health Economics 63*, 145–158.

Kirkeboen, L., E. Leuven, and M. Mogstad (2016). Field of Study, Earnings, and Self-Selection. *The Quarterly Journal of Economics 131*(3), 1057–1111.

Kirkeboen, L., E. Leuven, and M. Mogstad (2022). College as a Marriage Market. *NBER Working Paper 28688*, 1–52.

Lalive, R. and M. A. Cattaneo (2009). Social Interactions and Schooling Decisions. *The Review of Economics and Statistics 91*(3), 457–477.

Lee, D. S., J. McCrary, M. J. Moreira, and J. Porter (2022). Valid t-Ratio Inference for IV. *American Economic Review 112*(10), 3260–3290.

Leuven, E., H. Oosterbeek, and I. De Wolf (2013). The Effects of Medical School on Health Outcomes: Evidence from Admission Lotteries. *Journal of Health Economics 32*(4), 698–707.

List, J. A., F. Momeni, and Y. Zenou (2020). The Social Side of Early Human Capital Formation: Using a Field Experiment to Estimate the Causal Impact of Neighborhoods. *NBER Working Paper 28283*, 1–60.

Matta, J. and A. Orellana (2022). Neighbors' Spillovers on High School Choice. *Working Paper*, 1–59.

McHale, S., K. Updegraff, and S. Whiteman (2012). Sibling Relationships and Influences in Childhood and Adolescence. *Journal of Marriage and Family 74*(5), 913–930.

Nicoletti, C. and B. Rabe (2019). Sibling Spillover Effects in School Achievement. *Journal of Applied Econometrics 34*(4), 482–501.

Oettinger, G. S. (2000). Sibling Similarity in High School Graduation Outcomes: Causal Interdependency or Unobserved Heterogeneity? *Southern Economic Journal 66*(3), 631–648.

Qureshi, J. A. (2018a). Additional Returns to Investing in Girls' Education: Impact on Younger Sibling Human Capital. *The Economic Journal 128*(616), 3285–3319.

Qureshi, J. A. (2018b). Siblings, Teachers, and Spillovers on Academic Achievement. *Journal of Human Resources 53*(1), 272–297.

Raaum, O., K. G. Salvanes, and E. Sørensen (2003). The Impact of a Primary School Reform on Educational Stratification: A Norwegian Study of Neighbour and School Mate Correlations. *Swedish Economic Policy Review 10*, 143–170.

Słoczyński, T., S. D. Uysal, and J. M. Wooldridge (2022). Abadie's Kappa and Weighting Estimators of the Local Average Treatment Effect. *NBER Working Paper 15241*, 1–40.

van der Vleuten, M., J. Weesie, and I. Maas (2020). Sibling Influence in Field of Study Choices. *Research in Social Stratification and Mobility 68*(100525), 1–18.

Whiteman, S. D., S. M. McHale, and A. Soli (2011). Theoretical Perspectives on Sibling Relationships. *Journal of Family Theory and Review 3*(2), 124–139.

Wiswall, M. and B. Zafar (2015). Determinants of College Major Choice: Identification using an Information Experiment. *The Review of Economic Studies 82*(2), 791–824.

Zang, E., P. L. Tan, and P. J. Cook (2023). Sibling Spillovers: Having an Academically Successful Older Sibling May Be More Important for Children in Disadvantaged Families. *American Journal of Sociology 128*(5), 1529–1571.

Appendix

Table A1: Spillover effects excluding medicine and excluding all health-related fields

	Younger peer enrolls in any higher education study		
	Sibling	**Cousin**	**Neighbor**
Panel A: All lottery fields of study			
Older peer enrolls	0.057***	0.022***	0.003***
	(0.005)	(0.002)	(0.001)
Control complier mean	0.029	0.011	0.010
N	33,094	79,704	317,217
Panel B: Medicine excluded			
Older peer enrolls	0.031***	0.007***	0.001
	(0.005)	(0.001)	(0.001)
Control complier mean	0.010	0.003	0.002
N	12,274	32,681	114,415
Panel C: All health-related fields excluded			
Older peer enrolls	0.045***	0.010***	0.004
	(0.006)	(0.002)	(0.004)
Control complier mean	0.008	0.001	0.007
N	6,354	17,152	26,129

Notes: All regressions include controls for sex, non-western origin, age of the older peer at the first lottery, and fixed effects for lottery field of study, lottery category, year of the first lottery participation, and their interactions. All regressions use the normalized inverse of the propensity score as a weight. Standard errors are clustered at the older peer level. *p<0.1; **p<0.05; ***p<0.01

Table A2: Spillover effect of older peer's lottery result on younger peer's enrollment in any higher education study

	Younger peer enrolls in any higher education study		
	Sibling	**Cousin**	**Neighbor**
Panel A: Main results			
Older peer wins the first lottery	0.004	0.010*	0.000
	(0.006)	(0.005)	(0.004)
Control complier mean	0.797	0.654	0.489
N	33,094	79,704	317,217
Panel B: Heterogeneity by lottery category			
Older peer wins the first lottery	0.004	0.011*	0.007
	(0.007)	(0.006)	(0.004)
Older peer wins	0.002	-0.004	-0.019**
x 𝟙 [Older peer's GPA < 6.5]	(0.012)	(0.011)	(0.007)
N	33,094	79,704	317,217
Panel C: Heterogeneity by sex composition			
Older peer wins the first lottery	0.016**	0.009	-0.003
	(0.007)	(0.006)	(0.004)
Older peer wins	-0.023**	0.001	0.007*
x 𝟙 [Same sex = 1]	(0.010)	(0.007)	(0.004)
N	33,094	79,704	317,217
Panel D: Heterogeneity by age gap			
Older peer wins the first lottery	0.006	0.008	0.000
	(0.006)	(0.008)	(0.004)
Older peer wins	-0.008	0.001	0.005
x 𝟙 [Age gap > 5 (2)]	(0.011)	(0.009)	(0.004)
N	33,094	79,704	1,472,910
Panel E: Heterogeneity by parental income			
Older peer wins the first lottery	0.010	0.013*	-0.001
	(0.008)	(0.007)	(0.005)
Older peer wins	-0.013	-0.001	0.001
x 𝟙 [Parental income > median]	(0.011)	(0.010)	(0.007)
N	32,995	79,484	316,141

Notes: All regressions include controls for sex, non-western origin, age of the older peer at the first lottery, and fixed effects for lottery field of study, lottery category, year of the first lottery participation, and their interactions. All regressions use the normalized inverse of the propensity score as a weight. In Panel D, for siblings and cousins we include interactions with an indicator equal 1 if the age difference is larger than 5, and for neighbors – larger than 2. In Column (3) of Panel D, we include neighbors with at most 10 years gap. In Panel E, we include interactions with an indicator equal 1 if the parental income of older peers is above the sample median. Standard errors are clustered at the older peer level. *p<0.1; **p<0.05; ***p<0.01

Table A3: Robustness checks: different sample definitions

Panel A: Siblings	Oldest OS	Youngest OS	One OS
YS enrolls in the lottery	0.057***	0.057***	0.053***
field of study	(0.005)	(0.005)	(0.005)
YS enrolls in any	0.004	0.006	0.005
higher education study	(0.006)	(0.005)	(0.006)
N	33,094	33,094	31,990
Panel B: Cousins	Oldest OC	Youngest OC	One OC
YC enrolls in the lottery	0.022***	0.021***	0.021***
field of study	(0.002)	(0.002)	(0.002)
YC enrolls in any	0.010*	0.006	0.007
higher education study	(0.005)	(0.005)	(0.005)
N	79,704	79,704	70,798
Panel C: Neighbors	Youngest ON	At least one ON with $0 < p < 1$	At least one ON with $0.1 < p < 0.9$
YN enrolls in the lottery	0.003***	0.004***	0.004***
field of study	(0.001)	(0.002)	(0.001)
YN enrolls in any	0.000	-0.005	-0.004
higher education study	(0.004)	(0.004)	(0.004)
N	317,217	731,862	532,338

Notes: OS/YS - older/younger sibling, OC/OC - older/younger cousin, ON/YN - older/younger neighbor. All regressions use the normalized inverse of the propensity score as a weight. p is the propensity score measuring that at least one older neighbor is a lottery winner. In Panels A and B, standard errors are clustered at the older sibling/cousin level. In Column (1) of Panel C, standard errors are clustered at the older neighbor level, and in Columns (2) and (3) - at the neighborhood level. In all columns of Panels A and B and in Column (1) of Panel C, regressions include controls for sex, non-western origin, age of the older peer at the first lottery, and fixed effects for lottery field of study, lottery category, year of the first lottery participation, and their interactions. In Columns (2) and (3) of Panel C, regressions include the propensity score as a fixed effect. *p<0.1; **p<0.05; ***p<0.01