

Backes, Annika; Kovač, Dejan

Working Paper

Neighbor effects on human capital accumulation through college major choices

IWH Discussion Papers, No. 10/2025

Provided in Cooperation with:

Halle Institute for Economic Research (IWH) – Member of the Leibniz Association

Suggested Citation: Backes, Annika; Kovač, Dejan (2025) : Neighbor effects on human capital accumulation through college major choices, IWH Discussion Papers, No. 10/2025, Halle Institute for Economic Research (IWH), Halle (Saale), <https://doi.org/10.18717/dpf3jd-xk39>

This Version is available at:

<https://hdl.handle.net/10419/319913>

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.

Discussion Papers

No. 10

June 2025



Neighbor Effects on Human Capital Accumulation Through College Major Choices

Annika Backes, Dejan Kovač

<https://doi.org/10.18717/dpf3jd-xk39>

Authors

Annika Backes

Halle Institute for Economic Research (IWH) –
Member of the Leibniz Association,
Department of Structural Change and
Productivity
E-mail: annika.backes@iwh-halle.de

Dejan Kovač

Harvard University, Center for International
Development
E-mail: dejankovac@hks.harvard.edu

The responsibility for discussion papers lies solely with the individual authors. The views expressed herein do not necessarily represent those of IWH. The papers represent preliminary work and are circulated to encourage discussion with the authors. Citation of the discussion papers should account for their provisional character; a revised version may be available directly from the authors.

Comments and suggestions on the methods and results presented are welcome.

IWH Discussion Papers are indexed in RePEc-EconPapers and in ECONIS.

Editor

Halle Institute for Economic Research (IWH) –
Member of the Leibniz Association

Address: Kleine Maerkerstrasse 8
D-06108 Halle (Saale), Germany
Postal Address: P.O. Box 11 03 61
D-06017 Halle (Saale), Germany

Tel +49 345 7753 60
Fax +49 345 7753 820

www.iwh-halle.de

ISSN 2194-2188

Neighbor Effects on Human Capital Accumulation Through College Major Choices*

Abstract

Using the universe of high school and college admissions data in Croatia, we geocoded nearly half a million students' residential addresses to investigate how their college and major choices are influenced by older neighbors and peers. Using an RDD to exploit time and program variation in admission cutoffs, we find that having an older neighbor who was admitted to and enrolled in a program increases a student's probability of applying to the program by about 20%. We find that this effect consistently holds only for the closest neighbors, both in terms of distance and age difference. Female students are more likely to be influenced by older neighbors' choices, and male older neighbors' admission has a larger impact on both male and female students compared to female older neighbors. The effect is stronger if the student-neighbor pair lives in a region that does not have its own university, implying that the value of information in rural areas is higher. We find evidence that students don't follow their older neighbors to less competitive programs; instead, they are more likely to apply for the same programs their older neighbors were admitted to when the program is more prestigious. Next, we utilize the variation in weight scheme of Croatia's college study programs to show evidence, beyond college choices, of how older neighbors affect the human capital formation of their younger peers. The main channel through which we observe this effect is during high school, through specialization in the subjects needed to gain admittance to older neighbors' college programs. These findings shed light on the intricate dynamics shaping educational decisions and underscores the significant role older neighbors play in guiding younger peers toward specific academic pathways.

Keywords: college-major choice, human capital accumulation, neighbors, peer effects

JEL classification: D83, I23, I24, O15, R23

* We thank Joshua Angrist, Sarah Fritz, Stepan Jurajda, Asim Khwaja, Eliana La Ferrara, Steffen Müller, Johanna Raith, Simon Wiederhold and the seminar and conference participants at Halle Institute for Economic Research (IWH), Princeton University, Harvard University, AEDE conference, TriEcon Workshop and CEP Education Conference for helpful comments. For providing valuable support in data cleaning and management, we want to thank our research assistant Ivan Franic. For special help around data logistics from Agency for Science and Higher Education we thank Igor Drvodelic and Mirko Stanic.

1 Introduction

Choosing a field of study and college is a key life decision that influences one’s lifelong earnings trajectory. The share of individuals going to university is quite unequally distributed and is lower among disadvantaged students. High-achieving students who are low income are less likely to opt for ambitious education paths, despite the high returns of education (Hoxby and Avery 2013). Even among those students who decide to apply for college, the likelihood of whether they will apply to prestigious colleges and renowned study programs differs along the distribution of socioeconomic background. It does not only matter *if* you study, but also *what* and *where* you study, as there is such a large variation in long-run outcomes, such as earnings, both between top and less attractive universities as well as between study fields and major choices (Altonji et al. 2012; Altonji et al. 2016; Kirkeboen et al. 2016). Part of this mismatch can be attributed to unequal starting points for children, in terms of both institutional settings and the quality of information available within their close networks.

Equality of opportunity is a key topic in the literature on education. We can agree that individuals from different neighborhoods and backgrounds should have access to equal education opportunities; but in many countries and regions, even across the developed world, this is still not the case. It matters where you come from, even if, in theory, the options are the same. It has been shown that exposure to better neighborhoods increases college attendance, among the set of other long-run outcomes (Chetty et al. 2016). Still, the underlying mechanisms of this effect are unclear, due a lack of quality data and identification strategies. Historically, following (Angrist 2014), peer effects have been difficult to identify causally. For a better understanding of these mechanisms, it is crucial to disentangle neighborhood-level institutional factors, such as school quality, from neighborhood factors that occur at the individual level, such as peers. An important question arises: What part of the variation is explained by institutional factors, such as access to better education facilities, and what part is due to better peer effects? Our focus is on the latter.

In this study, we investigate neighbor spillovers on specific college \times major choices. Using a regression discontinuity design to estimate neighbor effects on college major choices, we provide causal evidence that older neighbors’ enrollment in a specific program has a spillover effect on younger students’ decisions. We use a unique micro dataset, taken from a centralized system that contains information on Croatian students’ high school history and college applications.

We find that older neighbors’ enrollment in a particular program increases the younger neighbor’s probability, evaluated at the sample mean, to apply for the same program—as their first preference by 23%, as one of their top three by 13-15%, and to apply to the

program at all by 14%. We further disentangle the possible mechanisms of the effect, first finding that younger female students seem to be influenced more strongly than male students by an older neighbor’s enrollment in a program, and that older male neighbors seem to have a stronger influence than female ones on younger students. Next, we find that students who live in cities without a university see a 37% increase in the probability, evaluated at the sample mean, that they will apply for a program their older neighbor is enrolled in, while for students living near a university there is no significant effect. This finding suggests a higher value of access to information in rural areas; whether this is through an aspiration effect or a purer information channel is still an open question. Finally, we find that younger neighbors follow their older neighbors to elite programs or programs with the highest enrollment requirements. The last two findings suggest that the scarcity of information in rural areas matters, and that it matters even more if students can observe an older neighbor who has managed to enroll in an elite program.

Next, we utilize the variation in study program weight schemes across Croatia’s college system to identify the causal effect of older neighbors’ admission to college programs on their younger neighbors’ human capital formation. College study programs in Croatia place a lot of weight on applicants’ grades in subjects connected to the field of study; for instance, a physics study program would prioritize applicants’ high school and state exam grades in physics. We use a difference-in-difference approach, using highly weighted high school subjects for the older neighbors’ admitted program as a treatment group and no-weight high school subjects as a control group, to identify the causal effect. We find a strong and positive effect through numerous robustness checks. After an older neighbor peer gets admitted to a particular study program, younger neighbor peers tend to increase their grades in subjects needed to get into that particular college program relatively more than they do for the rest of the control group subjects. The effect becomes stronger as the share of the weight placed on that particular subject in admissions increases. These interesting results show not only the effects older neighbors’ college choices have on their younger peers; they also highlight heretofore unexplored mechanisms of the effect. We show that younger neighbor peers research application requirements and the importance of grades in relevant subjects in order to increase their probability of admission to older neighbors’ college programs. Furthermore, there is evidence of the aspiration effect for younger neighbors because they invest relatively more time in subjects which would increase their likelihood of admission to their older neighbors’ study program; however, it was not possible to disentangle the extent to which this information channel operates directly or indirectly.

Finally, we contribute to the literature by providing additional evidence of direct exposure between neighbor-peers. Given that big administrative datasets like this that

encompass entire countries can have data at a detailed level, the neighbor-peer effects are measured as a Euclidean distance between the place of residence of two neighbors; we do not have data on direct interactions between neighbor-peer pairs. We bridge that gap by providing an additional measure of direct exposure between peers through social groups. We use a unique dataset on a universe of in-school and extracurricular activities for all students in Croatia for the observed period, through which we are able to identify neighbor peer pairs who also participated in the same social activities. Although we find a significant *negative* effect for sports activities, overall we find no enhancing effect of additional neighbor-peer interaction through different social activities. We highlight our data limitations on social activities in large administrative datasets and lay ground for future work on direct neighbor-peer interactions.

From a policy perspective, these findings can shed light on human capital formation and how it is impacted by the college choices older neighbors have. In particular, they highlight the importance of sharing admissions information in rural areas, which are usually underrepresented in college graduation rates, especially in elite programs. To our knowledge, this is the first study to investigate individual non-family spillover effects on specific college major choices and human capital formation in a large-scale setting, while providing additional evidence on the effects of direct neighbor-peer interactions and the main mechanism of this effect—younger students’ efforts to learn more about the subject weights and grades required to gain admission.

2 Literature Review

The returns of pursuing an education can vary widely depending on the context (Ashenfelter and Mooney 1968; Angrist and Krueger 1991; Angrist and Krueger 1992; Angrist and Krueger 1995; Angrist and Krueger 1994). In today’s globalized world, earning a college diploma can have a profound influence on an individual’s welfare. Interventions and mentoring can improve one’s college applications and attendance levels down the road (Carrell and Sacerdote 2017; Peter and Zambre 2017), and can also change one’s likelihood to focus on a certain field of study (Breda et al. 2021). In contrast, one might choose a less ambitious education path if they have asymmetric information, underestimate the effects of a diploma on their own future labor market value, or have lower aspirations due to a lack of role models in their social network. But apart from the positive impact on the individual, educational interventions can have positive spillovers on others in an individual’s family or social network as well (Carlana et al. 2022). To better target policies, then, it is beneficial to understand how and where these positive spillover effects work. We show that neighbor spillover effects on specific study program choices indeed exist. Despite being less pronounced than sibling spillovers, they do matter for certain

subgroups.

Our study contributes to several strands of literature. Primarily, it contributes to an increasing literature on how the quality of neighborhoods shape children’s educational outcomes (Brooks-Gunn et al. 1993; Ellen and Turner 1997; Jencks, Mayer, et al. 1990; Abdulkadiroğlu et al. 2014). Neighborhoods matter for several reasons. Chyn and Katz (2021) provide empirical evidence of the strong correlation of neighborhood quality and different labor market and educational outcomes for adults and children and discuss the mechanisms by which adults and children might be influenced in different neighborhoods. For children, the institutional setting of a neighborhood in regard to schooling possibilities, especially in regard to the quality of schools, can provide vastly different opportunities for their educational trajectory. Quality schools not only contribute to an increase in educational attainment, but they can offer students higher improvement in non-cognitive skills, as well as better teachers and mentoring resources. Apart from these mechanisms, which work mainly through the institutional settings of schools, another mechanism that might influence educational trajectories is the quality of peers, both in school and in one’s neighborhood.

Even with overwhelming evidence of neighborhood effects, there is still the question of how to causally interpret these findings. A set of recent experimental and quasi-experimental studies provides further evidence of the impact of better neighborhoods on children. In their seminal work, Chetty et al. (2016) exploit data from the Moving to Opportunity (MTO) experiment, where randomly selected families were offered vouchers to move to lower-poverty neighborhoods. They find positive effects of moving to a better neighborhood on students’ likelihood of attending college and on future earnings for children under the age of 13. For adolescents, they show small negative effects, which they explain through disruption effects. Nakamura et al. (2022) find further evidence for the positive impact of moving for young people on lifetime earnings and education in line with Chetty et al. (2016), using data on families’ forced migration after a volcanic eruption in Iceland. In a quasi-experimental setting using data on families moving across commuting zones and counties, Chetty and Hendren (2018a) and Chetty and Hendren (2018b) show that the length of exposure to a better neighborhood matters for children’s long-term outcomes. They find significant effects on earnings, college attendance, and other family-related patterns increasing with each year of exposure. We contribute to this literature by showing that part of the neighborhood effect on higher education outcomes can indeed be attributed to individual neighbor peer effects.

Our study also contributes to the literature on peer effects on education. The main challenge in identifying peer effects is the causal interpretation of correlated behavior in groups (Manski 1993; Manski 1995). The majority of the literature studies peer effects

on education choices and outcomes solely through family and sibling spillovers (Altmejd et al. 2021; Joensen and Nielsen 2018; Qureshi 2018). There are two main reasons for a such a methodological approach. The first is that we are certain that siblings interact, so the exposure and connection is already identified. The second is that the direction of causality is clear: One can assume that, in terms of educational choices and spillovers, older siblings influence younger ones. However, there is generally a lack of data to properly identify exposure to other non-family peers.

Estrada et al. (2021) investigate the impact of older cohorts' admission rates to elite schools on the application, admission, and school performance of younger cohorts in their respective home school. They find that older students' admission to elite schools increases the probability that younger peers will apply and get admitted to these schools among groups whose parents have a lower education background. However, the effect in this school admission system is hard to disentangle. Aggregated at school level, the result cannot be interpreted solely as a "peer effect." Other underlying channels might include social peer effects, quality of teacher effects, and general school effects. Using data from an admissions lottery for oversubscribed study programs in the Netherlands, Avdeev et al. (2024) explore educational spillovers of siblings, cousins, and neighbors on their younger counterparts. They find substantial family spillover effects and, to a lesser extent, neighbor spillovers. This is in contrast to the findings in Altmejd et al. (2021), where siblings do tend to follow the same college or college-major combination but not generally the same major; the authors explain the difference in findings due to small differences between the quality of universities in the Netherlands. Their setting differs from ours in that Dutch students use a field-of-study-wide lottery for admission, meaning that if a student does not get in, they have to choose a different field. In contrast, Croatia uses a program (college \times major) admission threshold, so we concentrate on these more granular choice-specific spillovers, which include choices on both field and place of study.

The closest to our study is Barrios-Fernández (2022). He finds that having a close neighbor who was admitted to a university and eligible for student loans increases students' probability of enrolling in a university by 10 percentage points. This effect decreases with distance. Unlike his setting, which uses a RDD around financial loan eligibility to assess impact on the closest neighbors, our study uses a RDD setting around college program quotas as an instrument for enrollment status. This distinguishes the mechanisms of the effects in our study; Barrios-Fernández (2022) sheds light on how financial aid information from older neighbors influences college enrollment, but in our setting the information channel is broader, because attending public colleges in Croatia is free. Thus, the information we observe might be more related to things that vary by program—for instance, quality of teachers, study plan or curriculum, or simply broadening the set of

opportunities which prior didn't seem feasible.

3 Data

Our data consists of four main parts: college, high school, neighbor, and social activities data.

The college data is taken from the official registry of the Agency for Science and Higher Education in Croatia.¹ It contains the universe of students applying to colleges from 2010 to 2015. Every private or public college/university in Croatia is included in this system. For every applicant, we observe the list of college programs, ranking of every program, and which college program they ultimately enrolled in. For every school year, we also observe program quotas, which allows us to calculate year-program specific cutoffs.

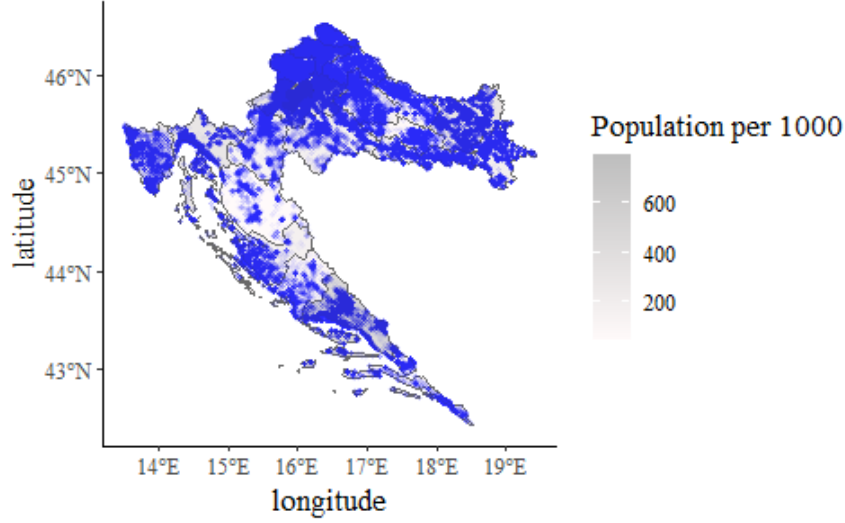
The high school and elementary school data is taken from the official registry of the Ministry of Science in Croatia. It contains the universe of high school and elementary grades for all students in Croatia from 2008 to 2015. For every student and for every school year, we observe: residence address, grades for every subject, yearly GPA, and number of justified and unjustified absences. We also observe the identity of the head teacher for every classroom and school/classroom identifiers.

To obtain neighbor network data, we geocoded close to 500,000 residence addresses of individuals, see Figure 1. For every individual, we identified the three closest neighbors, via Euclidean distance, out of the pool of addresses of individuals that attended the same high school. We only include student-neighbor pairs where the distance between homes is below 10km. Besides distance in meters, we also divided neighbors into three additional groups, from one to three years difference in school years. In the end, our matrix of neighbors has two dimensions: distance between residence and years difference in school cohorts. This creates a 3x3 matrix, which we use in different specifications throughout the paper. The reason for using such a matrix is to test whether the magnitude of the effect changes with the increase of neighbor distance or age difference. Figure 2 shows the distribution of distance for the first, second, and third neighbor of the student neighbor pairs that are one year apart. Most neighbors live in within 2,000m proximity, and the first, second, and third neighbors live an average of 1.07km, 1.57km and 1.90km apart, respectively.

Finally, we obtain data on social activities from the official registry of the Ministry of Science in Croatia as well, for the same time span of the educational data, to show us information on within and outside-of-school activities for the universe of students.

¹Administrative data from the central applications office (NISpVU) and the Agency for Science and Higher Education (ASHE).

Figure 1: Geo-coded Residential Addresses of Students (2010-2015)



Note: X-axis shows the longitude, y-axis shows latitude. Blue dots are residential addresses of students. Grey areas show county level population density per 1000.

4 Empirical Strategy

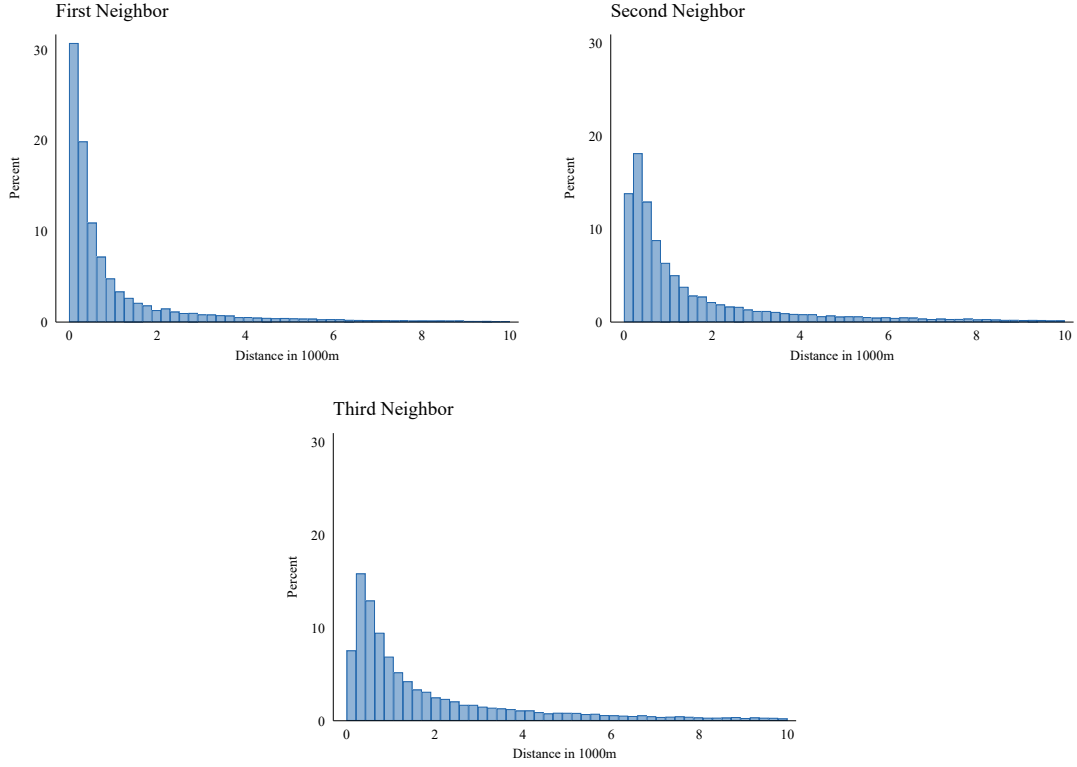
In this section, we provide the historical setting underlying the specification of our regression discontinuity design and placebo tests.

4.1 The Centralized Application System

The centralized college application in Croatia was created in 2010. The main goals of moving from a decentralized to a centralized system were to increase transparency, remove information asymmetries across different universities, and offer applicants real-time information on their enrollment status. Prior to 2010, the biggest problem with Croatia's higher education system was the lack of transparency in college admissions. Students could compete for several positions at several universities, and in the end, could only enroll in one program. This created large problems with under-supply of some and over-supply of other set of programs.

The centralized application system is an online platform offering each applicant a unique portal. All relevant program information is easily accessible: quota, program information, tuition (if applicable), points system, etc. Each applicant can choose up to 10 programs from the hundreds offered in a given school year. They list their selected programs in preference order from 1st to 10th. Based on their high school grades, state exam points, international competition, disability status, etc., they receive program specific points, a weighted sum of the previously mentioned components. One applicant might

Figure 2: Neighbor Distance 1 year gap



Note: X-axis shows the distance between student and older neighbor in 1000m, y-axis shows the percentage share of observations per distance bin.

have different points for different programs based on the weights each program have decided to assign, as universities have complete autonomy in deciding weight structures for their programs.

For the observed time period, 2010 to 2015, the average cohort size was between 35,000 and 45,000 applicants, and the number of available college programs ranged from 700 to 900. The system had a 4- to 5-day window in which the final points rankings are announced. The rankings update every hour, and during these 4 to 5 days, applicants can change their selected college programs or how they rank them. The system is a repeated deferred acceptance algorithm, where individuals are only accepted to their highest ranked preference for which they had more points than the program cutoff points.

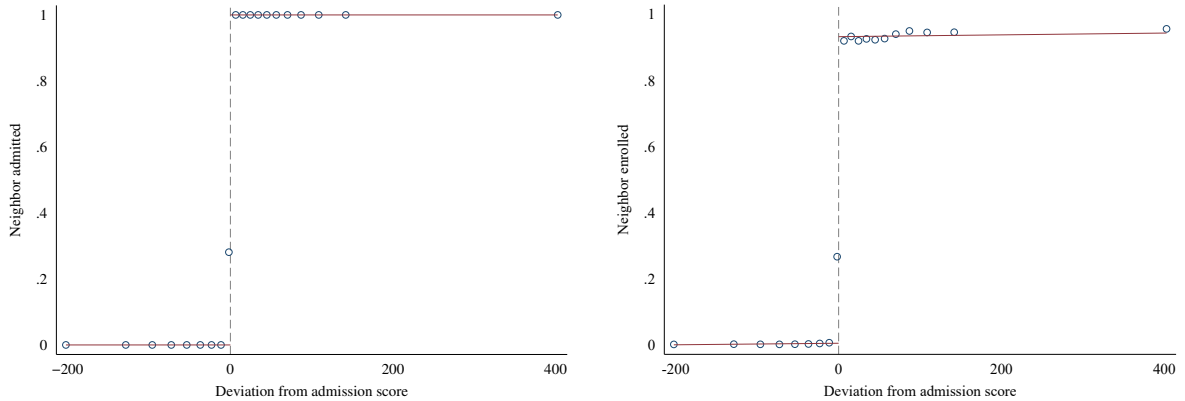
Final assignments cannot be changed; students have an option to decline their college program match, but that would mean they need to wait until the next year to apply and, at that point, would need to pay full tuition. The magnitude of costs they would incur for tuition was from 3 to 5 average Croatian yearly salaries, so they are incentivized to accept their assignment on the first attempt.

4.2 Regression Discontinuity Design

We use the universe of high school and college admission data in Croatia from 2010 to 2015, exploiting time and program variation in college cutoffs to create an RDD that examines causal estimates: comparing the impact on college-major choices of students where older neighbors got into a program to those whose older neighbors nearly missed the quota (following Altmejd et al. 2021). The centralized college admission system in Croatia has about 700 to 800 programs and quotas per year, and about 75% of programs end up oversubscribed, which allows us to identify older neighbor-students around the cutoff.

We identify neighbors by calculating Euclidean distances between students' home addresses. For every student, we calculate the distance to closest three neighbors attending the same high school, in order to observe more variety in exposure to older neighbors. Our aim is also to test whether having a more distant neighbor implies having less exposure to him/her.

Figure 3: Admission score, admission probability and enrollment probability



Note: X-axis shows the deviation from the admission score needed to get admitted to the program. Y-axis on the left: student got admitted =1, Y-axis on the right: Student enrolled after admission =1.

Figure 3 shows the admission and enrollment probabilities of older neighbors in our dataset. Students apply to up to 10 programs and must rank their choices. They will only get admitted to one program, and admissions rankings for each program are strictly computed according to high school grades and centralized state exam results. The final number of points needed to get admitted to a certain program—i.e., the cutoff point—is unknown by students during the application process. Students above the cutoff have a 100% probability of getting admitted to a given program and enroll with a probability of roughly 95%. As implied earlier, most students accept their assigned study program, as they lose their admission fee waiver otherwise.

Table 1: Student Characteristics, Full Sample and Bandwidth Samples: Above and Below the Cutoff

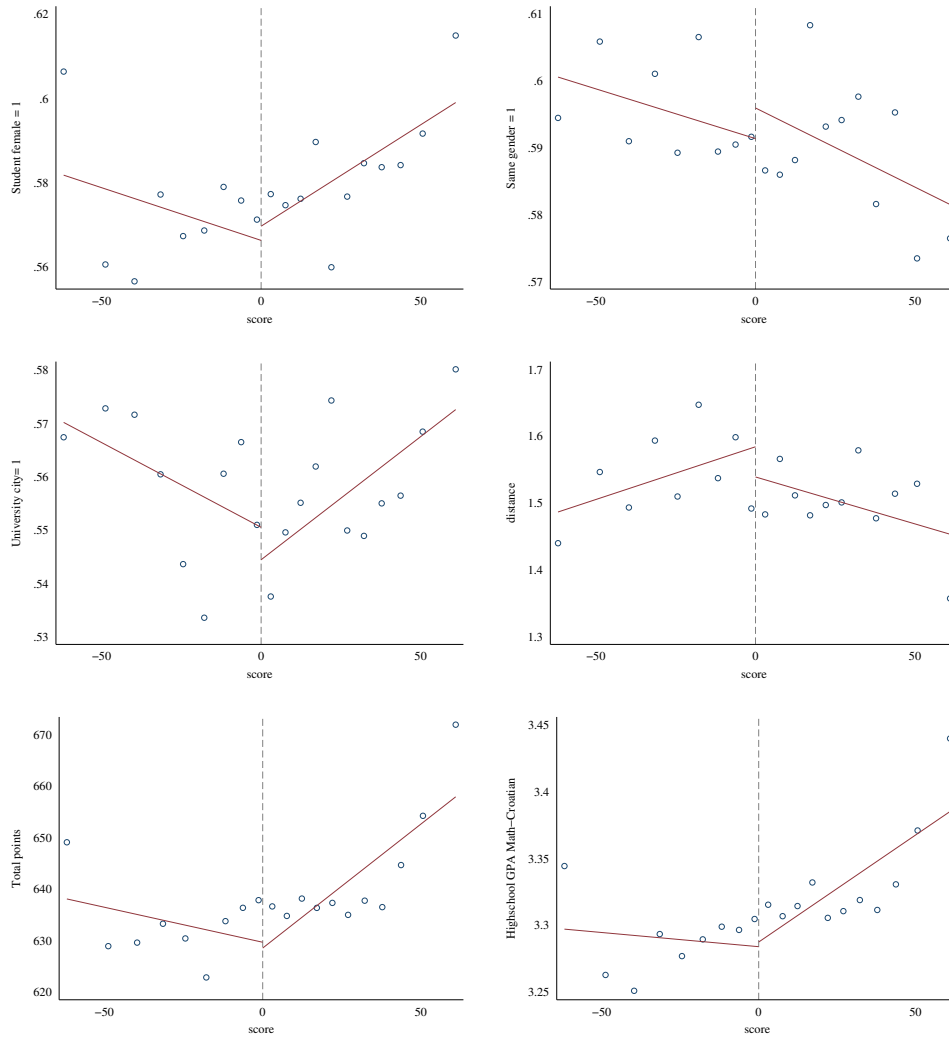
	(1) All		(2) Above		(3) Below	
	mean	sd	mean	sd	mean	sd
Panel (a): Full sample						
Female Students	0.571	0.495	0.573	0.495	0.568	0.495
Female Neighbors	0.608	0.488	0.588	0.492	0.635	0.482
Student lives in city with University/College	0.555	0.497	0.554	0.497	0.556	0.497
Student-Neighbor pairs has same gender	0.591	0.492	0.584	0.493	0.601	0.490
Distance in 1000m	1.544	1.987	1.493	1.952	1.609	2.030
Total_points	635.693	239.010	647.004	252.583	621.061	219.352
Observations	421235		237575		183660	
	mean	sd	mean	sd	mean	sd
Panel (b): 50 points +/- cutoff						
Female Students	0.574	0.494	0.579	0.494	0.568	0.495
Female Neighbors	0.627	0.484	0.622	0.485	0.634	0.482
Student lives in city with University/College	0.554	0.497	0.553	0.497	0.556	0.497
Student-Neighbor pairs has same gender	0.594	0.491	0.593	0.491	0.595	0.491
Distance in 1000m	1.532	1.977	1.518	1.958	1.551	2.002
Total_points	634.390	223.434	636.187	217.469	631.848	231.461
Observations	195950		114162		81815	
	mean	sd	mean	sd	mean	sd
Panel (c): 10 % of points +/- the cutoff						
Female Students	0.579	0.494	0.583	0.493	0.573	0.495
Female Neighbors	0.638	0.481	0.631	0.482	0.647	0.478
Student lives in city with University/College	0.558	0.497	0.557	0.497	0.559	0.496
Student-Neighbor pairs has same gender	0.592	0.491	0.589	0.492	0.595	0.491
Distance in 1000m	1.518	1.970	1.499	1.950	1.541	1.994
Total_points	638.255	232.073	642.065	229.641	633.420	235.034
Observations	226329		126582		99747	

Note: Sub-samples above and below the admission cutoff for different bandwidths are shown. Includes the full sample for 1st to 3rd neighbor and age difference of 1-3 years.

In the full sample, we include student-neighbor pairs with an age difference of one to three years and identify the first, second, and third closest neighbor of each student for each year gap. Table 1 shows the descriptive statistics of the main characteristics of students in the dataset. Panel (a) shows student characteristics for the full sample, panel (b) and (c) show the sub-samples where the older neighbor's application score was 50 points or 10% of the total score below or above the admission score cutoff. For each panel, column (1) shows descriptive statistics for all observations in the sub-sample and columns (2) and (3) show the same but limited to above or below the admission cutoff respectively. Our sample consists of 57.1 % female students and 60.8% female neighbors.

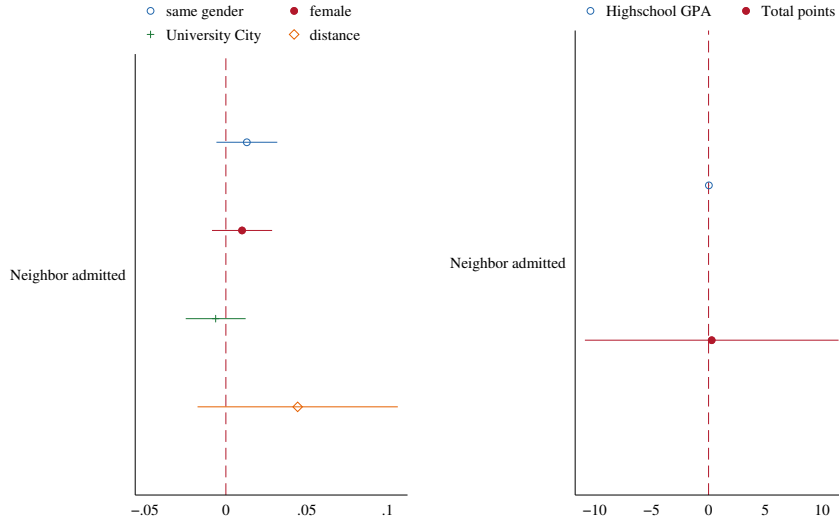
55.5 % of these student-neighbor pairs live in a city or region with direct university or college access at the time of applying to university. 59.1 % of the sample consists of pairs of the same gender. The average distance between the younger students' and older neighbors' home addresses is 1.54 km. Our distance measure is broader than that of Barrios-Fernández (2022), where the average distance to the closest neighbor is 0.09 km. Due to challenges during the geocoding process for rural regions, that study concentrated on the three largest metropolitan areas in Chile, while our study concentrates on the universe of students throughout Croatia.

Figure 4: Student Characteristics



Note: Graphs show binned scatter plots for the first neighbor 1 year gap sample, only including observations where the application score is 10% above or below the cutoff. X-axis shows the deviation from the admission score required for admission to the program. Y-axis on the left depicts the respective independent variable of the student. Plots from left to right display in the first row the student gender (female==1) and gender composition of the student-neighbor pair (=1 if of the same gender), second row University city status (=1 if the pair lives in a city that has direct college/university access and distance (in km). In the third row plots for total points of the younger student and GPA for math and Croatian GPA re shown.

Figure 5: Student Characteristics - Coefficients

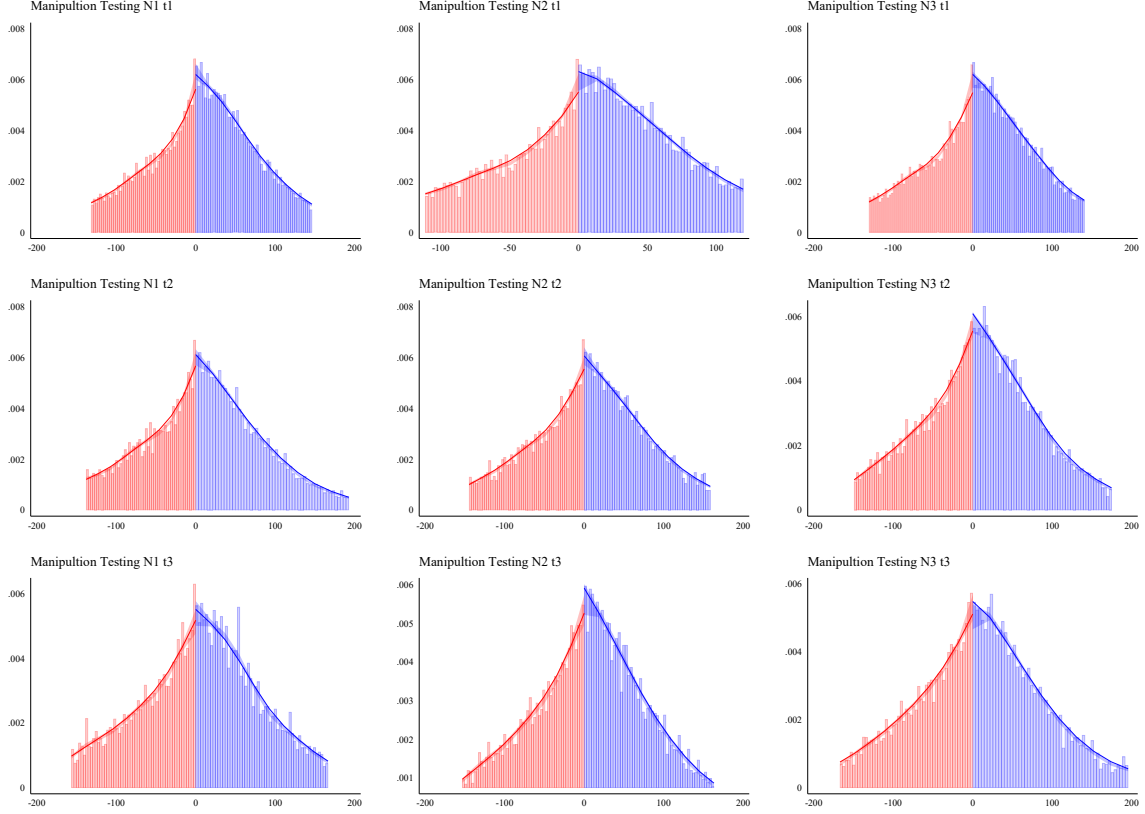


Note: These graphs display coefficients of regressions testing for significant jumps around the cutoff for variables plotted in Figure 4. The left graph shows coefficients for the probability of student-neighbor pairs to be of the same gender, the student's gender, the student living in a city with direct university access, and the distance between neighbors. The right graph shows the coefficient testing for difference between total application points and high school GPA above and below the admission cutoff.

When comparing the sub-samples above and below the admission cutoff, we do not see any significant differences between these characteristics, except for the difference in total application points of younger students around the cutoff. We therefore investigate possible jumps of the independent variables closer to the cutoff. Figure 4 shows binned scatter plots for student characteristics. We do not see any systematic jumps. Figure 5 shows coefficients when testing for the significance of possible jumps in a regression setting. Estimations follow the strategy described in Section 4.3 when using student characteristics as the dependent variable instead of the outcome of interest, students applying to the same program.

Our main outcome of interest is to see whether the younger student applies for the same college \times major combination (program) as the older neighbor as a) their first preference, b) their first to third preference, or c) any preference. Most students, around 85%, get an offer from one of their first three preferences. That is why we specifically focus on the results for dependent variable b), applying to their neighbor's first to third preference.

Figure 6: Density plots for neighbours across residence distance and age difference



Notes: Density plots for 3 x 3 matrix - first to third neighbor and 1 to 3 year gap sub-samples

Finally, in our specifications we focus on different subsets of neighbors, either by distance in residence or distance in age cohorts. For distance in residence, we focus on the three closest neighbors, and for the age difference, on groups from one to three years apart. Density plots 6 shows no systematic jumps at 0.

4.3 Econometric Specification

To investigate the reduced form effect of an older neighbor's admission to a program on the preference set of a younger student, following Altmejd et al. (2021), we run regressions as specified by

$$\rho_{icmt} = \beta \times \text{abovecutoff}_{icmt-n} + f(\alpha_{icmt-n}; \theta) + \mu_{cmt-n} + \epsilon_{icmt} \quad (1)$$

where ρ_{icmt} signifies whether a student from student-neighbor pair i applies to the same college c and major m choice in application year t as his or her neighbor in $t-n$. We include college \times major fixed effect μ and thereby only compare students application probabilities whose neighbors applied for the same program. ϵ_{icmt} is the error term. $\text{abovecutoff}_{icmt-n}$ is a binary variable equal to 1 if the neighbor's application score α_{icm} in year $t-n$ is

larger or equal to the necessary admission score s for the college \times major combination cm in the same year, and 0 otherwise.

$$\text{abovecutoff}_{icmt-n} = \begin{cases} 1, & \alpha_{icmt-n} \geq s_{cmt-n} \\ 0, & \alpha_{icmt-n} < s_{cmt-n} \end{cases} \quad (2)$$

$f(\alpha_{icmt-n}; \theta)$ is a linear function of the application score of the older neighbor of the applicant-neighbor pair i for major-college choice in year $t - n$:

$$f(\alpha_{icmt-n}; \theta) = \theta_0 \alpha_{icmt-n} + \theta_1 \alpha_{icmt-n} \times 1[\alpha_{icmt-n} > s_{icmt-n}] \quad (3)$$

which allows different slopes on both sides of the admission cutoff. To identify the effect of enrollment, we instrument the older student's enrollment ($\text{enrollment}_{imct-n}$) with the admission outcome ($\text{abovecutoff}_{icmt-n}$) of the college \times major choice in a fuzzy RD setting.

4.4 Placebo Test

We also conducted a placebo test. The direction of causality in our identification strategy is clear: college choices of older neighbors affect college choices of younger peers. In order to placebo-test our hypothesis, we reversed the order— i.e., we tested whether the college choices of younger neighbors affect the college choices of older neighbors. In all of our specifications, we find no significant estimates. Table 2 shows detailed results with regressions including control variables for all outcome variables for the first neighbor and one-year age gap. Corresponding tables for the second and third neighbor are shown in the Appendix A, Tables 7 and 8. For an consistent overview, Figure 7 shows coefficients of testing whether younger students' enrollment had an effect on older neighbors' choices for the first, second, and third closest neighbors for all three variants of the dependent variable investigated, with and without control variables. All coefficients are consistently statistically insignificant for all specifications using both the data-driven bandwidth (columns 2 and 5) and the bandwidth relative to the admission score (columns 3 and 6).

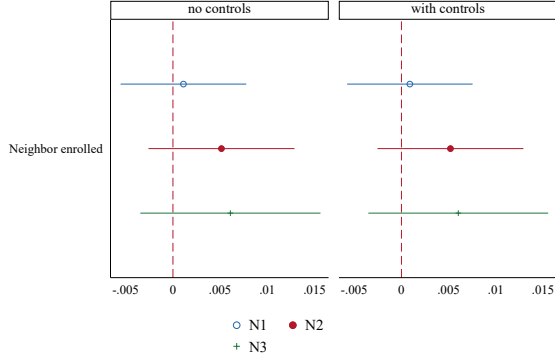
Table 2: Placebo test: Closest Neighbor and 1 Year Age Gap

	Neighbor admitted			Neighbor enrolled (fuzzy)		
	(1)	(2)	(3)	(4)	(5)	(6)
	Rdrobust	OLS BW select	OLS BW 10%	Rdrobust	IV BW select	IV
Panel (a): Applying same program as 1st choice						
Robust	0.0062 (0.004)			0.0068 (0.004)		
Neighbor admitted		0.0025 (0.003)	0.0008 (0.003)			
Neighbor enrolled					0.0024 (0.004)	0.0009 (0.003)
Student controls	YES	YES	YES	YES	YES	YES
fstage					61981	75354
ymean	0.03	0.03	0.03	0.03	0.03	0.03
BW	55.72	55.72		50.97	50.97	
N	59861	32152	34040	59861	29784	34040
Panel (b): Applying same program as 1st - 3rd choice						
Robust	0.0121** (0.006)			0.0130** (0.007)		
Neighbor admitted		0.0064 (0.005)	0.0042 (0.005)			
Neighbor enrolled					0.0060 (0.006)	0.0046 (0.005)
Student controls	YES	YES	YES	YES	YES	YES
fstage					62797	75354
ymean	0.07	0.06	0.06	0.07	0.06	0.06
BW	63.89	63.89		52.20	52.20	
N	59861	35582	34040	59861	30389	34040
Panel (c): Applying same program any choice						
Robust	0.0144** (0.007)			0.0155** (0.007)		
Neighbor admitted		0.0082 (0.006)	0.0040 (0.006)			
Neighbor enrolled					0.0079 (0.006)	0.0044 (0.006)
Student controls	YES	YES	YES	YES	YES	YES
BW	73.01	73.01		70.96	70.96	
fstage					79816	75354
ymean	0.12	0.11	0.11	0.12	0.11	0.11
N	59861	37957	34040	59861	37377	34040

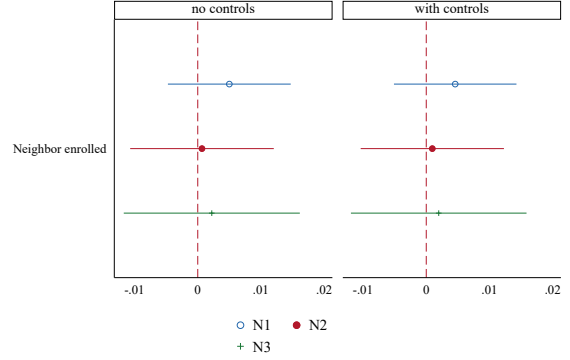
Notes: Placebo test examines the effect of a younger student's admission to and enrollment in a program on older neighbor's choice set. Columns (1)-(3) show the reduced-form results for the student's admission, and columns (4)-(6) for the student's enrollment. Panels (a)-(c) show results for dependent variables choosing the same program as 1st preference, 1st-3rd preference, or any preference. Columns (1) and (4) show results for the rdrobust estimation and the resulting bandwidths. Column (2) and (5) show results of the linear regression using the data driven bandwidths from the preceding model, Columns (3) and (6) the results when using 10% bandwidths relative to the total admission score per program. All regressions include college \times major and year fixed effects. Standard errors are clustered at student-neighbor pair level. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Figure 7: Placebo N1 to N3

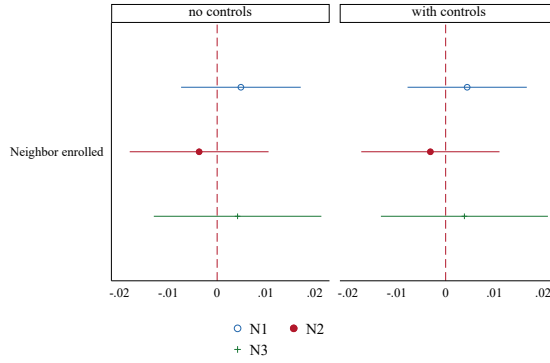
Applying 1st preference, 10 % bandwidth



Applying 1st-3rd preference, 10 % bandwidth



Applying any preference, 10 % bandwidth



Note: Graphs coefficients of the Placebo regressions (2SLS), 10% bandwidth, N1 to N3 for 1 year gap, with and without control variables

5 Results

In this section, we provide three sets of results: neighbor proximity effects, the effect on human capital formation with difference-in-difference approach, and neighbor peer effects in social groups. The neighbor proximity effect is our basic RDD, well established in the literature and similar to Altmejd et al. (2021) and Barrios-Fernández (2022). The second part uses a novel difference-in-difference approach to estimate the neighbor proximity effect through older neighbors' college choice on the human capital formation of younger peers. Finally, our study is the first study to provide evidence of neighbor-peer interaction alongside measurements of proximity of residence. Although we use the universe of students participating in different types of social activities, we cannot neglect a selection bias toward social activities in general, and different types of social activities. Thus, we provide the RDD results for a sample of neighbor peer pairs who also participated in the same social activities.

5.1 Neighbor Proximity Effects

Figure 8 shows binned scatter plots with the mean of the respective variable conditional on the deviation of older neighbors' application scores from the admission scores and 95% confidence bands. For each of the variables, we see a significant jump in the probability of applying for the same program if an older neighbor got accepted to the program. Evaluated at the sample mean, this gap is most considerable for the probability of applying to that program as a first preference. This is descriptive evidence using the full sample, for all neighbors and all year gaps, and not controlling for student characteristics. We see that the share of students applying for the same program is constant if the older neighbor had an application score below the admission cutoff, and is significantly higher if the older neighbor got admitted. For student-neighbor pairs where the older neighbor got admitted to a program, the probability that the younger student applies to the same program increases with the application score.

In our main analysis, we concentrate on the sub-sample of student-neighbor pairs of only the closest neighbor and one cohort ($T - 1$ to T) apart.² Table 3 shows results for the reduced form (effect of admission, columns 1-3) and for the effect of the older neighbor's enrollment (columns 4-6). For each model, we start by running a regression with `rdrobust` (column 1 and 4) to identify an optimal data-driven bandwidth using a common MSE-optimal bandwidth selector for the RD treatment effect estimator (following Calonico et al. 2014). We then run, for each specification, the linear regression using the data-driven bandwidth and an additional bandwidth relative to the admission score (10% of the admission score left and right of the cutoff), reported in columns 2-3 and 5-6. Panel (a) reports results on the younger student's application decision for the first preference, Panel (b) for the first to third preference, and Panel (c) for applying to any preference. The effect of the older neighbor's enrollment is consistently larger. Having a one-year-older closest neighbor who enrolled in a program increases the younger students' probability of applying for the same program as their first preference by 0.7 percentage points, applying as one of the first three preferences by 0.8–1.0 percentage points, and applying at all by 1.3–1.5 percentage points. Evaluated at the sample mean, the probability of applying for the same program increases by 23% as a first preference, by 13–15% as a top-three preference, and 14% for applying to the program at all.³

²We run several specifications also for the second and third closest neighbors and $(T - 2)$, $(T - 3)$ but only find persistent and significant effects for the closest neighbors and $(T - 1)$. Graphs showing resulting coefficients of the IV estimations are shown in the Appendix, Figure 17 and 18

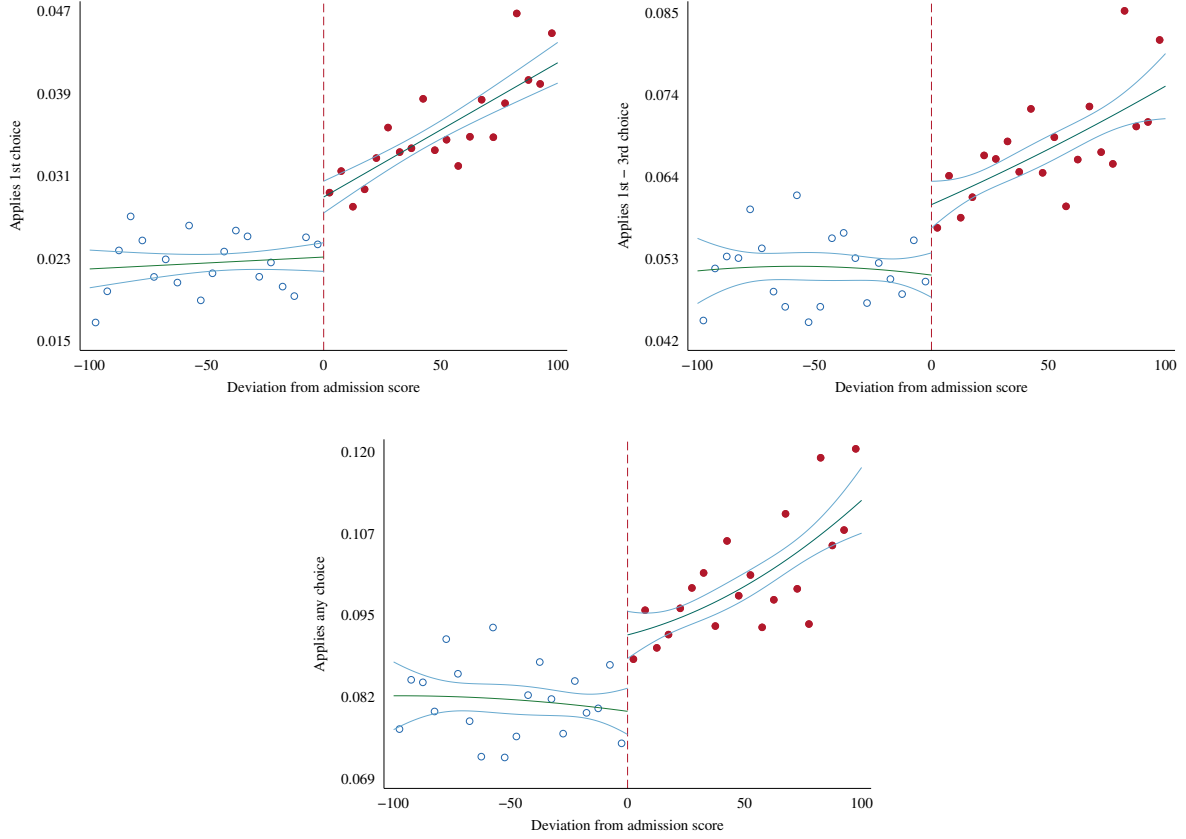
³We also run the same regressions including the set of student characteristics as controls and results are robust, the magnitude of the coefficients only varies maximum by 0.01.

Table 3: Main Results: Closest Neighbor and 1 Year Age Gap

	Neighbor admitted			Neighbor enrolled (fuzzy)		
	(1)	(2)	(3)	(4)	(5)	(6)
	Rdrobust	OLS BW select	OLS BW 10%	Rdrobust	IV BW select	IV
Panel (a): Applying for the same program as 1st choice						
Robust	0.0115*** (0.004)			0.0110** (0.005)		
Neighbor admitted		0.0060* (0.003)	0.0066** (0.003)			
Neighbor enrolled					0.0072* (0.004)	0.0073** (0.004)
BW	70.92	70.92		57.83	57.83	
ymean	0.035	0.032	0.030	0.035	0.031	0.030
fstage					59170	67529
N	65383	39598	35725	65383	34318	35725
Panel (b): Applying for the same program as 1st - 3rd choice						
Robust	0.0163*** (0.005)			0.0161** (0.006)		
Neighbor admitted		0.0082** (0.004)	0.0077* (0.005)			
Neighbor enrolled					0.0101* (0.005)	0.0085* (0.005)
BW	86.85	86.85		57.25	57.25	
ymean	0.070	0.066	0.063	0.070	0.065	0.063
fstage					58822	67529
N	65383	45041	35725	65383	34006	35725
Panel (c): Applying for the same program, any choice (1-10)						
Robust	0.0251*** (0.007)			0.0281*** (0.008)		
Neighbor admitted		0.0126** (0.005)	0.0113** (0.006)			
Neighbor enrolled					0.0135** (0.007)	0.0125** (0.006)
BW	76.75	76.75		55.43	55.43	
ymean	0.105	0.099	0.095	0.105	0.097	0.095
fstage					57311	67529
N	65383	41799	35725	65383	33182	35725

Notes: Columns (1)-(3) show the reduced form results for the neighbor's admission, columns (4)-(6) for the neighbor's enrollment. Panels (a)-(c) show results for dependent variables choosing the same program as 1st preference, as 1st-3rd preference, or any preference. Columns (1) and (4) show results for the rdrobust estimation and the resulting bandwidths. Column (2) and (5) show results of the linear regression using the data driven bandwidths from the preceding model, Columns (3) and (6) the results when using 10% bandwidths relative to the total admission score per program. All regressions include college \times major and year fixed effects. Standard errors are clustered at student-neighbor pair level. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Figure 8: Students Following Older Neighbors

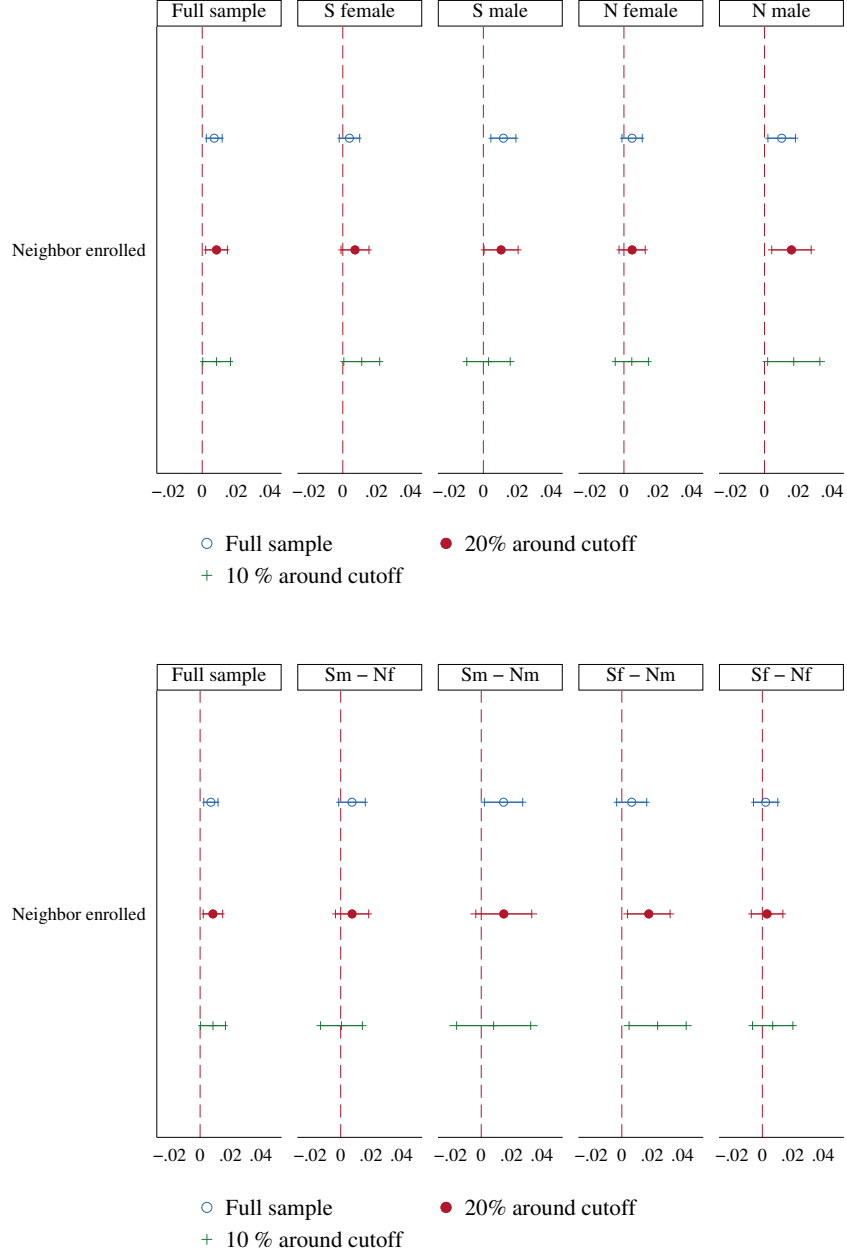


Note: Graphs show means and 95% confidence bands of dependent variables conditional on the deviation of the neighbor's application score from the admission score (x axis). Y axis of the upper left graph depicts the share of younger students applying for the same college-major combination as the older neighbor as a first preference; the upper right graph shows the same for the younger student applying to the same program as first to third preference; and the lower graph shows the same for the younger student applying to the same program as any preference.

5.1.1 Gender Differences

In the next step, we investigate which subgroups drive these results. As most students get accepted to one of the first three preferences they note in the application process, we focus on the results that show the effect of older neighbors enrollment in the program on the younger student's probability to apply to that program as their first, second, or third preference. Figure 9 shows results separately for student and neighbor gender and by gender groups. The left graph shows results for sub-samples where either the student or neighbor are female or male and displays coefficients for the full sample and using a 10% or 20 % bandwidth, respectively. We concentrate on results for the 10% bandwidth, where older neighbors with application scores just left and right of the cutoff are most similar.

Figure 9: Results by gender



Notes: Results by different gender groups are shown. Displayed are coefficients from the 2SLS specification of the first neighbor-one year gap sample using different the full sample and a 20% and 10 % bandwidth. All regressions include college \times major and year fixed effects. Standard errors are clustered at student-neighbor pair level.

Younger female students seem to be influenced more strongly than male students by an older neighbor's enrollment, and older male neighbors seem to have a stronger influence than older female ones. This is also seen when comparing specific gender groups; only the sub group with female students and male neighbors shows a significant effect on

application probability.⁴

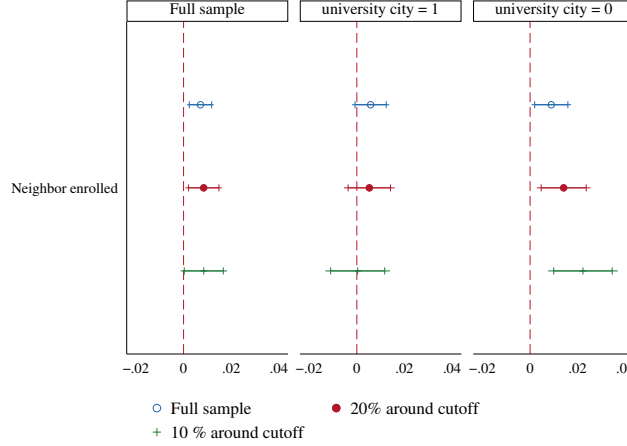
5.1.2 Rural and Urban Divide

In order to investigate the value of information, we now look at differences in regard to proximity to a university or college. We identify the students that (during high school) live in a city or region with direct university access. This can be important for two reasons: Living in a region with direct university access is also closely correlated with having more information about those schools in near proximity; and universities are usually situated in places with a higher population density. One might have visited a nearby college before or know many people connected to it. Having one close neighbor who went to this school might not have as significant an impact, because of the abundance of connections and valuable information. Additionally, for students with strong geographical preferences to stay close to home, having a university close by would make it much more likely that they will choose a program there rather than at any other school, independent of their neighbors' choices. Figure 10 shows results for students living in a region with a university (`has_uni=1`) and without.⁵ Students living in places nearby universities and colleges do not follow their closest older neighbors to the same programs, but students with no direct university access, where the value of information is higher, are more likely to choose the same program as their neighbors. For those students, having an older neighbor enroll in a specific program increases their probability of applying for that program by 2.2 percentage points, which is a 37% increase compared to the sample mean (of this sub-sample).

⁴Interestingly, these subgroup results differ quite severely when looking at the probability to apply for first preference or any preference. For the male neighbor- female student pair, the effect on first preference is very small, but quite considerable on applying for any preference. For the female neighbor-female student pair we only find a significant effect on applying as first preference, but none on the other outcomes. One possible explanation here is that part of the male neighbor-female student pair effect might be driven by a "marriage market effect" based on romantic connections, where in most cases the male part is the older one of the pair. For the female neighbor - female student pair, having an older female role model might increase the probability to study exactly the same program and not just any program which would be geographically closer.

⁵We identify this variable by matching the place were the student lives when applying to college with all possible cities that have a university or college. `has_uni = 1` for all students where the place of living matches one of these cities.

Figure 10: Results proximity to college

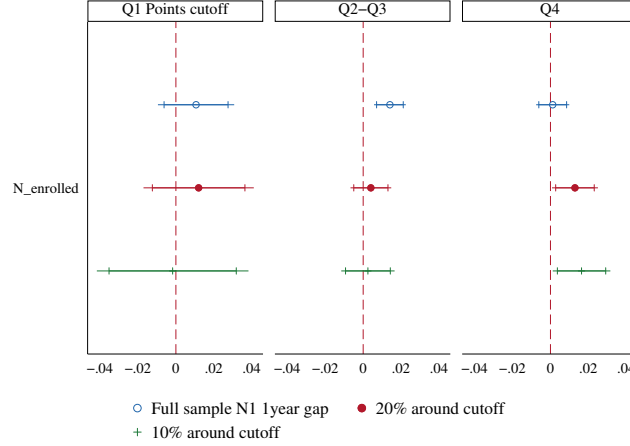


Notes: Results by difference to university and college proximity are shown, where $has_uni = 1$ implies that at the time of applying the student lives in a regions with a university or college nearby. Displayed are coefficients from the 2SLS specification of the first neighbor–one year gap sample using: the full sample and a 20% and 10 % bandwidth. All regressions include college \times major and year fixed effects. Standard errors are clustered at student-neighbor pair level.

5.1.3 Elite and Non-Elite College Programs

Another point to consider is to which programs students actually follow their neighbors to. This leads to the question of whether following a neighbor to study the exact same thing actually seems reasonable. If a student applies to a program solely because their neighbor did, irrespective of their own skill sets or other intrinsic motivation, it would hardly seem rational. We therefore identify program groups depending on the admission score required to get admitted. Programs with high admission scores are of high demand, which serves as a proxy for "elite" programs of high quality. Figure 11 shows results by program demand: Q1 represents observations for low-demand programs, where access and admission is easy. Q2-Q3 are medium-demand programs and Q4 high-demand programs. The total effect of the full sample seems to be fully driven by high-demand programs. This can be interpreted as an aspiration effect, where students follow their older neighbors specifically to elite programs because they might otherwise not have thought it worthwhile to try. Older neighbors might also share valuable insights for the application process afterward, which might encourage younger students to apply. Or older neighbors' might have positive experiences with the program, leading younger students to consider programs they otherwise wouldn't have.

Figure 11: Results by Program Demand



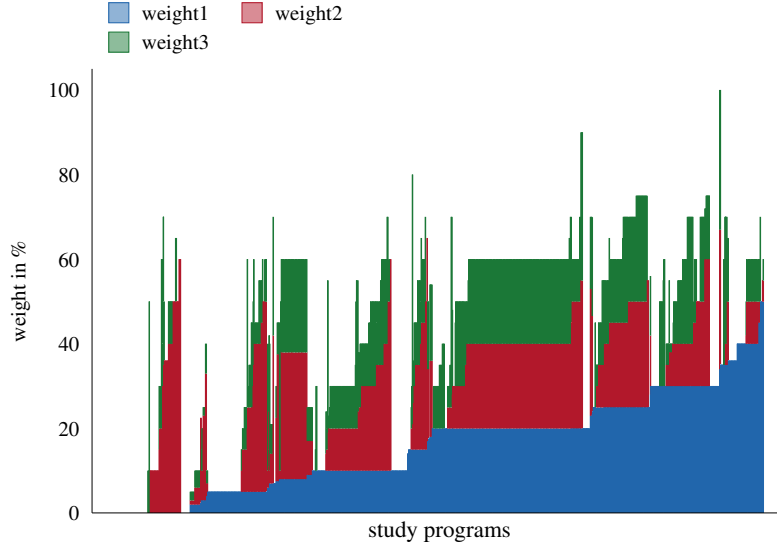
Notes: Results by difference in program demand are shown. Programs in Q1 are low-demand programs, Q4 high-demand programs. Displayed are coefficients from the 2SLS specification of the first neighbor-one year gap sample using different the full sample and a 20% and 10 % bandwidth. All regressions include college \times major and year fixed effects. Standard errors are clustered at student-neighbor pair level.

5.2 Neighbor Effects on Human Capital Formation

Besides having an effect on college choices, in this section we explore whether older neighbor peers have an effect on the human capital formation of younger peers. We exploit the variation in weight schemes across study programs in Croatia to investigate the important mechanisms of this effect. Are younger neighbor peers only affected by the choice of their older neighbors, or do they also internalize the strategies needed to get into a particular study program?

In Croatia, universities have the autonomy to decide how to weight the various requirements of their study programs. Figure 12 shows the variation in these weights by programs for an average school year. The main subjects that contribute to the weight scheme the most remain stable across years. As visible in Figure 12, the autonomy of decision-making produces the variation that might shape different application strategies across individuals, but also educational performance at earlier stages. In this section, we focus on the latter.

Figure 12: Variation in Program Weights



Notes: X axis shows 900 program ids, y axis shows the stacked weights by program id

Total application scores by program are the result of a weighting scheme of grades from subjects in high school, overall GPA in high school, state exam percentage by subject, winning of international competitions, and in rare cases, an entry exam. Figure 12 shows, by study program, the weights that the state exam results per subject have on the total score. In the majority of cases, the weight scheme of a particular study programs is connected to the subjects needed for successful completion of that program. For instance, in order to study physics as a study program, the biggest weight is on the high school or state exam grades from physics or other related subjects, such as math. It is visible in Figure 12, in blue, that some weights make up up to 60% of the total requirements—i.e., focusing on just component of the weight scheme can dramatically increase one’s likelihood of admission to that program. A detailed information set about the college weight scheme and high school subjects used in our analysis is in Appendix B.

We use a difference-in-difference approach to investigate whether an older neighbor’s admission to college, besides encouraging younger neighbors to follow them to the same study programs, might also have an effect on the educational performance of younger neighbor peers. We focus only on three set of neighbor-student pairs: for all admitted older neighbors, 10 and 5 percent bandwidths around the cutoff, as in the previous section. The question here is: If the students are following their older neighbors in terms of college choices, are they also adjusting their educational performance to increase their likelihood of admission in the future? To test this, we divide the set of 700 programs by dispersion of weights: programs placing a big weight on particular subject/s and programs versus those having more evenly distributed weights. Next, for each neighbor-student pair, we

focus on the study program to which the older neighbor was accepted. The set of high school subjects that are included in the weights or are mandatory to pass for successful application are our treatment group, while the subjects not included in the weight scheme are our control group—i.e., if the younger neighbor peers are indeed internalizing the college choices of their older peers, they would potentially also at that point learn more about the subjects needed to get into that particular study program. Whether they learn directly from their older peers or the older peers’ admission to that particular study program motivates them to learn more themselves remains an open question. Our event is the timing of admission of marginally accepted older neighbors, which is orthogonal to the educational performance of students during high school. We focus only on the four grades of high school, from 9th to 12th, since there is a very low probability that students would think about their application to college during elementary school. If the older neighbor was accepted to a college program while the younger student was in the 10th grade, than we would expect to see the jump in grades in the preceding period, 11th to 12th grades, for subjects in our treatment group—i.e., the set of subjects that have a high share in the weight of a older neighbor’s study program. In this analysis, we focus only on closest older neighbors, $N=1$, who got admitted to a study program, not on $N=2$ and $N=3$, the next two in terms of distance proximity measure.

In Table 4, we show the difference-in-difference results for $N=1$ (i.e., the closest neighbor) and $T=2$ (i.e., two years difference between older and younger neighbor peers).⁶ This specification allows us to observe two periods before and after the older neighbor got accepted to a particular study program. In Appendix Tables 9 and 10, we show results for $T=1$ and $T=3$ specifications and elaborate on our results. In column 1, table 4, the average treatment effect is positive and significant. The difference in means of high school grades between the treatment and control subjects, before and after the older neighbor got accepted to a "treated" college program, is 0.048. Thus, within the period after their older neighbor got admitted to that particular study program, younger neighbor peers increase their grades relatively more for subjects needed to get into older neighbors’ study programs than for subjects not needed to gain admittance. For study programs where the aggregated weights of all state exam subjects is low, up to 30% of the total, in column 2 of table 4, we find no effect. As the weight of state exam subjects in a study program increases, column 3 of table 4, we find a significant and positive effect, almost two times

⁶For this analysis we exclude grades in both Croatian and Math, as these are the subjects that are included in most programs as either mandatory to pass or with weights. The fact that these subjects are important is therefore common knowledge and it is mandatory to take the state exams in these subjects, and would be falsely included in the treatment group for all students. For further information on subjects and weights, see Appendix B.

Table 4: DiD Results for closest neighbor (N=1) and two years gap (T=2)

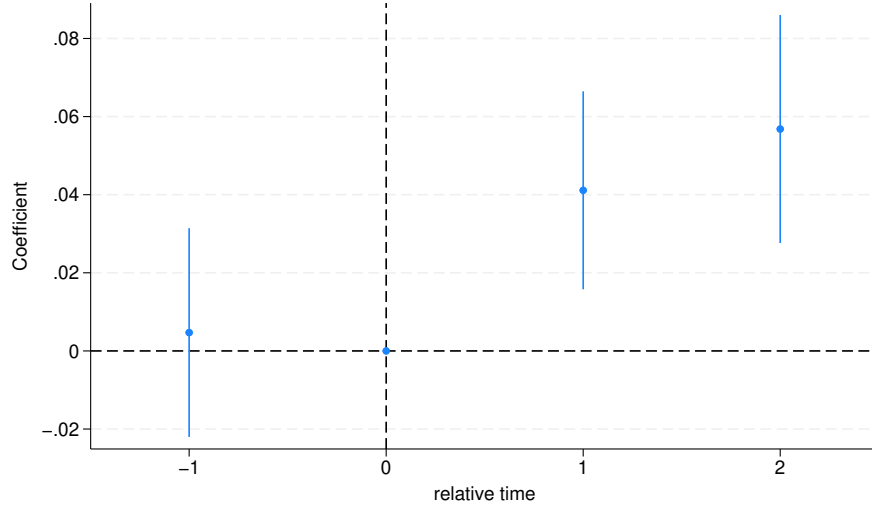
	(1)	(2)	(3)	(4)	(5)
	main	Q1 weights	Q4 weights	Q1 GPA	Q4 GPA
	b/se	b/se	b/se	b/se	b/se
timeT (post treatment = 1)	-0.0534*** (0.009)	-0.0318* (0.016)	-0.0705*** (0.020)	-0.0687*** (0.020)	-0.0211** (0.010)
subjT (treated subject = 1)	-0.0168 (0.013)	0.0055 (0.024)	-0.0601** (0.027)	-0.0301 (0.024)	-0.0214 (0.015)
timeT (=1) \times subjT (=1)	0.0485*** (0.012)	0.0252 (0.023)	0.0754*** (0.024)	0.0458** (0.021)	0.0465*** (0.014)
N	83404	24560	21680	20777	20541

Notes: Full sample for T2 is comparison for grades of subjects mandatory/ with weights of the program the older student got admitted to vs grades of other subjects, within student variation, Q1 means students where weights of the neighbor's program make up only 30 percent of the total final score, Q4 55 or more percent of final score. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

higher in the magnitude than the overall average treatment effect. Finally, in column 4 and 5 of table 4, we divide our sample into high- and low-GPA students by quartiles⁷. We find a similar effect for both low and high GPA students—i.e., it is not the case that low-GPA students, who have a bigger possibility for an increase, necessarily utilize it. Figure 13 depicts our results graphically by time periods before and after the older neighbor was admitted to a program. T=0 is the end of the younger peer's school year in which the older neighbor got admitted to a particular study program. The y axis are the plotted coefficients from Table 5 column 1. For T=2, we observe a significant and positive increase for grades in treatment subjects relative to control subjects.

⁷GPA calculated by student as the mean of all subjects included in the analysis over all 4 years of high school

Figure 13: Main specification (T=2) for closest neighbor (N=1)



Notes: DiD coefficients for the closest neighbor (N=1) with two years' difference in age. T=0 is the school year when older neighbor was admitted to a particular college program. College admissions are in July, while the school year ends in June, so this event happens a month after younger student has finished 10th grade.

Table 5: DiD coefficients by time periods for T=2

	(1) main b/se	(2) Q1 weights b/se	(3) Q4 weights b/se	(4) Q1 GPA b/se	(5) Q4 GPA b/se
t-1	0.0070 (0.008)	0.0223 (0.019)	0.0227 (0.016)	0.0189 (0.018)	-0.0129* (0.008)
t+1	0.0039 (0.008)	0.0124 (0.020)	-0.0140 (0.016)	-0.0359** (0.017)	0.0029 (0.008)
t+2	0.1246*** (0.011)	0.1257*** (0.036)	0.1206*** (0.021)	0.0393* (0.021)	0.0336*** (0.010)
subjT	-0.0182 (0.015)	0.0055 (0.027)	-0.0473 (0.030)	-0.0109 (0.027)	-0.0287 (0.018)
t-1 × subjT	0.0047 (0.014)	0.0022 (0.027)	-0.0252 (0.028)	-0.0389 (0.027)	0.0151 (0.020)
t+1 × subjT	0.0411*** (0.013)	0.0181 (0.026)	0.0679*** (0.025)	0.0352 (0.025)	0.0422** (0.018)
t+2 × subjT	0.0568*** (0.015)	0.0304 (0.029)	0.0570** (0.028)	0.0148 (0.028)	0.0631*** (0.019)
N	83404	24560	21680	20777	20541

Notes: full sample for T2 is comparison for grades of subjects mandatory/ with weights of the program the older student got admitted to vs. grades of other subjects, within student variation, Q1 means students where weights make up only 30 percent of the total final score, Q4 55 or more percent of final score, subjT= 1 means "treated" subjects important for older neighbors program, treatment occurs at ttt=1, all regressions include student fixed effects and are clustered at student x subjT level, including year and subject controls. * p < 0.1, ** p < 0.05, *** p < 0.01

Table 5 shows the coefficients by time period before and after the older neighbor got admitted to a study program, relative to the treatment year. In column 1 of table 5, the biggest relative increase of treatment subject grades to control subject grades was in the 12th grade. As students get closer to college, they increase their grades relatively more. In column 2, for low aggregated weights by program, we find no effect. In column 3, for aggregated weights bigger than 55%, we find the biggest increase in the 11th grade. Finally, in columns 4 and 5, we divide the sample by quartiles of younger peers' high school grades. We find, in column 5, that high-GPA students (Q4) have a positive and significant effect, while low-GPA students (Q1) do not have a significant effect.

In addition to $T=2$, the closest neighbor-pair year gap of two years, in the Appendix we provide results on the other two groups, $T=1$ and $T=3$, as well as for the closest neighbors. For $T=1$, a one-year difference, we find only a small effect on the increase of the treatment subject grades. One explanation might be that one year is a too short a period to improve in certain subjects. For $T=3$, the three year difference, we find a positive and significant effect only for 12th grade, but no aggregated effect. Similarly, for $T=2$ results, as the weights for "treatment subjects" increase, the magnitude of the effect doubles in the first year after the older neighbor gets admitted. It is important to note the difference in the interpretation of coefficients between $T=1$, $T=2$, and $T=3$. As the age gap increases, the tenure at the admitted study program of the older neighbor also increases. Comparing the estimated joint effect for $T=1$ and $T=3$, in terms of information about the study program experience, varies. Since we cannot distinguish whether this was a direct or indirect information effect, we cannot disentangle the overall effect. Figure 19 and Tables 9 and 10 provide a detailed set of results. Additionally, the timing of treatment makes it harder to interpret these results, as for $T=1$ there is only one post-treatment year and for $T=3$ there is only one pre-treatment year.

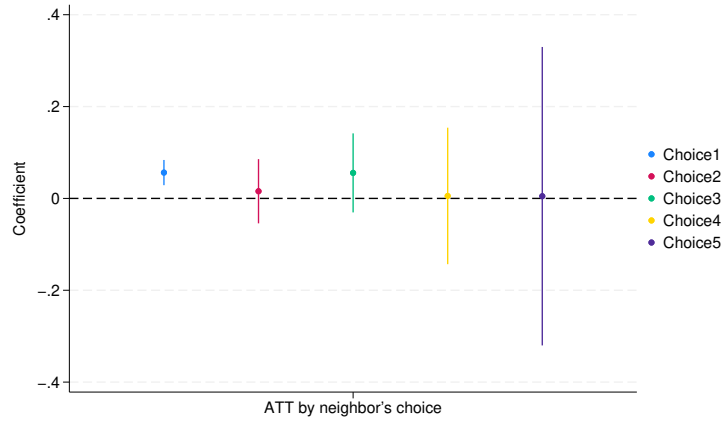
5.2.1 Robustness and Placebo Analysis

In this section, we provide a variety of robustness checks and placebo analyses, which largely confirm our initial findings. In the previous section, we provide evidence of the positive effect of older neighbors' college admission on younger students' human capital accumulation through specialization in grades. The effect varies depending on the age difference, having a strong effect only when the two have a two-year age difference, while effects are smaller or not significant for the one-year and two-year differences. Next, we look at the variety in the choice sets of both older neighbors and younger students to better understand what drives this effect.

We run separate regressions by the admitted choice rank of older neighbor. Every individual can rank up to ten choices on their rank order preference list, and after running

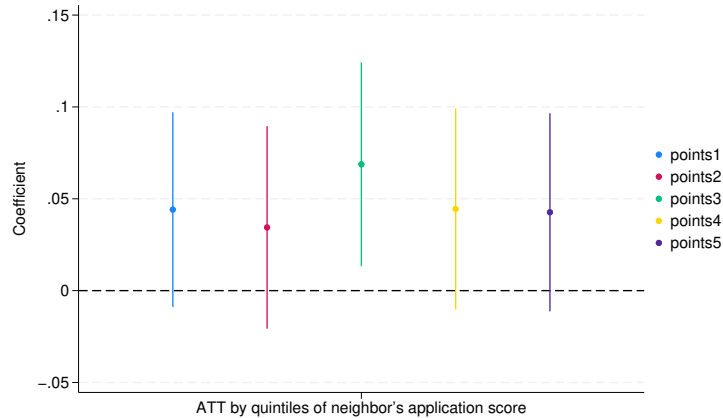
the deferred acceptance algorithm, the system commits them to enroll in their highest ranked choice. In Figure 14, for our main specifications, $N=1$ and $T=2$, we provide DiD estimates by choice rank. Only when neighbor got admitted to his/her first-ranked choice do we find a significant and positive effect. Next, in Figure 15, we divide the neighbors' admission scores into quantiles—i.e., their rank within their admitted college program. Interestingly, we find no significant effect when neighbors are at the top or at the bottom of the distribution, but rather when they are in the middle of the distribution.

Figure 14: DiD by first to fifth choice of neighbor for $N=1$ and $T=2$



Notes: The figure shows DiD coefficients by older neighbor's admitted choice rank for our main specification - closest neighbor ($N=1$) and two years apart ($T=2$). X axis are choices from 1st to 5th, Y axis is the estimated DiD coefficient for that particular subgroup.

Figure 15: N1T2 DiD by quantiles of neighbors admission score



Notes: The figure shows DiD coefficients by older neighbor's admitted choice rank for our main specification: closest neighbor ($N=1$) and two years apart ($T=2$). X axis are the quantiles of rank distribution to admitted college program of the older neighbor, Y axis is the estimated DiD coefficient for that particular subgroup.

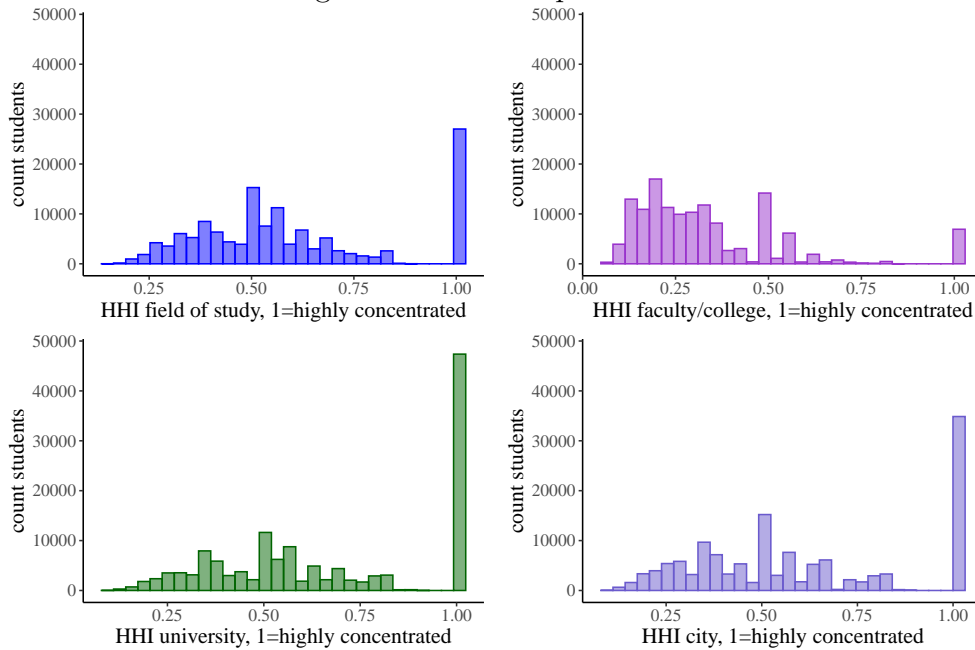
Finding oneself at the top of the rankings for a particular set of college programs

might not be unexpected. Older neighbors with extraordinary grades may have been on a path to that particular study program long before his/her admission date. Hence, to add an additional component of uncertainty, we focus on the bandwidth of older neighbors around the program cutoff—barely admitted vs. barely not admitted. For 5 and 10 percent bandwidths around a college program’s cutoff, Figures 21 and 22, we find similar patterns for closest admitted neighbors one and two years apart as we find in our main specification. For 3 years apart, Figure 23, due to decreasing number of observations for that particular group, we find no overall effect. Finally, for the group of not-admitted closest older neighbors, Figure 20, from one to three years apart, we find overall no effect. This last analysis serves as an additional placebo test.

To strengthen our argument, we need to rule out several additional possibilities. First is the fact that there might be an inflation trend in grades for the set of subjects for which the younger student would apply, irrespective of the older neighbor effect. Our several additional empirical analysis in the last paragraphs, by design, resolve this issue. Second, we want to rule out unobserved neighborhood-level elements that might be driving the results and be the cause of the preference homophily of neighbor-student pairs toward the same set of subjects. In some neighborhoods, there might be a permanent demand for a particular set of jobs that drives a selection into similar set of study programs—for instance, agricultural college programs in rural areas, or a particular set of industries in a certain urban area. To be able to distinguish the driving effects, we analyzed the set of preferences within each neighbor-student pair. Each student can rank up to 10 college choices in their list. We divided their choices into four main categories based on their preferences: field of study, college, university, and city. In this analysis, we wanted to distinguish that the student is really following their older neighbor and not only a set of study programs that may have similar weighing scheme in place. Each year, out of 700 to 800 programs, applicants can choose only 10 study programs; but there are clusters of programs, which are field-oriented, that have similar weight-scheme for particular set of subjects. For instance, getting admitted to physics across different colleges/universities may have similar weighing schemes. In order to have a different set of types of applicants, we used the entire dataset on college preferences and divided students by our four parameters. Some students may only express preferences for a particular field, while others may prefer to live in the same city irrespective of the field of study. For each of the four parameters, in Figure 16, we created a distribution of preference concentration using the Herfindahl–Hirschman Index (HHI). A value of 1 represents highly concentrated preferences (e.g., 1 field of study in choice set of 10), a value close to 0.1 shows mixed preferences (e.g., 0.1, meaning 10 different fields of study among 10 choices). The reason to compare the rank order lists of neighbors and students is because they do not know

which of the programs they will be admitted to. Every year, close to 40,000 students apply and rank their preferences. If students' preferences are mixed, including study programs that are not field oriented, then admission from one ranked preference to another would imply a large jump in the weights across subjects. In Figure 16, the largest number of individuals with mixed preferences—many with values below 0.5—are in the college group. The distribution of highly concentrated preferences (a value of 1) is greatest among universities and cities, indicating strong geographical preferences. For field of study, there is a substantial number of students with both low and high concentration values.

Figure 16: Student preferences



Note: The graphs show Herfindahl–Hirschman Index (HHI) distributions of students choice sets regarding field of study, colleges, universities and cities. A value of 1 represents highly concentrated preferences (e.g. 1 field of study in choice set 10), a value close to 0.1 shows very mixed preferences (e.g. 0.1 meaning 10 fields of study among all 10 choices). Excluded are all choice sets with only 1 choice reported.

In Table 13 in Appendix C, for our main specification, closest neighbor and two years apart, we divided the three parameters: city, field of study, and university by first and last quartile of distribution—i.e., comparing students who have clear preferences for a particular parameter and mixed preferences. In columns (1) and (2), for city/ geographical preferences, we find a positive and significant effect only for clear (highly concentrated) student choice sets and no effect for choice sets with lower concentration. In columns (3) and (4), for clear preferences and mixed preferences for field of study, we find an effect only for choice sets with a lower HHI. Finally, in columns (5) and (6), for university, we find a positive and significant effect for clear preferences for a particular university. The

fact that we do not find any effect for clear field preferences and find an effect for city and university suggests that younger students follow older neighbors even when there is a big difference in weight schemes across study programs.

Next, in Table 14 (Appendix C), similar to Table 13, we divide the sample according to the distribution of the older neighbor’s preference strength. We also compare the first and fourth quartile for city, field of study, and university choice concentration. We only find a positive and statistically significant effect for the subsets with clear neighbor preferences, and no effect for clear or mixed field of study choice sets.

This section provides evidence that older neighbor peers and their success getting admitted to college study programs has an effect on human capital formation of younger neighbor peers. The channel of human capital formation is through specialization in subjects that are needed for their future enrollment in the study program. We provide evidence that younger neighbor peers increase their grades in the subjects needed to enroll in an older neighbor’s program relatively more than for subjects not needed for the admitted program. These findings also provide evidence of an important mechanism of the effect—learning about the study program weights in order to maximize their own eventual likelihood of admission. Whether this is a direct effect of older neighbors providing information on study program weights, or an indirect effect of older peers motivating younger neighbor peers to find information on and subsequently study harder for higher weights subjects, remains an open question.

5.3 The Effect of Neighbor Peers in Social Groups

In this section, we utilize the universe of data on social activities, within and outside of school, for the same period as the rest of our educational outcomes. The neighbor effects in the literature so far were calculated through distance in residence between neighbor peer pairs without any evidence of direct interaction. The first part of our paper utilizes the same identification strategy, and we call it the "neighbor proximity" effect, which is different from this "direct neighbor" effect. We use social activities data to identify neighbor-student pairs who were: (1) closest neighbors; (2) went to the same high school, and (3) engaged in the exact same social groups for the same period. These novel estimates will shed light on the subset of neighbor-student pairs who were enrolled in the same high school and had, to some extent, additional interaction beyond living next door to one another. It is not a perfect measure of direct peer interaction, but it is a step forward and can provide additional evidence of possible interaction in different social activities. The choice of participating in activities in general, but also different types of social activities, is correlated with educational performance and other individual-level variables, so we extend our RDD strategy from the first part in order to remedy

any possible selection bias. Our empirical strategy still compares neighbor-student pairs around the quota cutoff, but also controls for the fact that they were engaged in the same social activities, within and outside of school. Since we are comparing neighbor student pairs within the same activity, by design, any selection bias when comparing pairs in the same social activity is largely ruled out in our regressions. For instance, we will be comparing closest neighbors, enrolled in the same school and participating in the same soccer or chess club.

Our analysis of social activities is very granular, but it is not without its limitations. For the original set of neighbor-student pairs, we analyzed all possible social activities in which they engaged for the observed period. The division of in- and outside-of-school activities is provided to us by the Ministry. An individual might be engaged both in-school and outside-of-school activities. For some activities, there are clear names, such as soccer club "FC Dinamo," while some in-school activities might be named more generally, like "soccer club." We distinguish those as separate, assigning the latter as soccer activity for that particular school. Each school offered a range of different activities, and we analyzed all of them. To minimize the range of our analysis, we cluster the same identifiers of social activities in the same groups. For instance, all social group identifiers in soccer clubs outside of school were in one group (FC Dinamo and FC Hajduk), while all social group identifiers of soccer across schools were in another group (soccer activities in two different schools). For all activities, we focused on the top 50 clusters of groups. As our procedure of finding student-neighbor pairs was already computationally heavy, we focused on the top 50 groups of activities is to increase the likelihood of finding a match.

Different social activities have different selection probabilities based on the observable variables. By comparing neighbor-student pairs enrolled in the same social activity around the older neighbor's program cutoff, we control for any potential selection biases that would arise if we compared pairs across different social activities.

In Table 6, we provide RDD results for our three main dependent variables: the probability that younger student is choosing the same college as a first preference, same college program as one of their top three, or choosing the older neighbor's admitted program as any of their top 10. In columns 1 to 3 we provide the conventional, bias-corrected, and robust RDD results for all of the prior selected top 50 clusters of activities. For the sample of neighbor-student pairs who went to the same high school and were enrolled in the same activity of any type, we find no effect. Due to the limited number of observations, we had to cluster our regression analysis into two large groups of activities that are similar in terms of the frequency of social meetings and the type. First group we name the "study" group. It primarily consists of in-school activities that aim to provide additional learning opportunities in different subjects. The group consists of the

Table 6: RDD Results for different sets of social activities

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	All activities	All activities	All activities	Study	Study	Study	Sports	Sports	Sports
	b/se	b/se	b/se	b/se	b/se	b/se	b/se	b/se	b/se
Conventional	-0.016 (0.02)	-0.030 (0.03)	0.010 (0.04)	0.084 (0.06)	0.103 (0.08)	0.133 (0.09)	-0.024* (0.01)	0.031 (0.08)	-0.222 (0.17)
Bias-corrected	-0.021 (0.02)	-0.036 (0.03)	0.006 (0.04)	0.096 (0.06)	0.071 (0.08)	0.061 (0.09)	-0.032** (0.01)	0.025 (0.08)	-0.269 (0.17)
Robust	-0.021 (0.02)	-0.036 (0.03)	0.006 (0.04)	0.096 (0.08)	0.071 (0.08)	0.061 (0.10)	-0.032* (0.02)	0.025 (0.10)	-0.269 (0.20)
Observations	2449	2449	2449	112	112	112	266	266	266

Note: RDD coefficients for three main outcome variables of younger students: probability of having the 1st, top 3, or any out of 10 preferences match the older neighbors' admitted college program. Columns (1)–(3) are for all activities. Columns (4)–(6) are for the group of study activities. Columns (7)–(9) are for the group of sports activities. . * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

following social activities: book club, geography, history, mathematics, ecology, arts, and German and English language. In columns 4 to 6 we find a positive, but not significant, effect. We also clustered the different sports activities, both within and outside of school, into a "sports" group. The sport group consists of the following activities: school-level soccer, handball and basketball; and sport club activities in soccer, handball, basketball, volleyball and water-polo. Thus, both school soccer club and soccer club outside of the school are in this group, but the neighbor and student pair would have to be enrolled in the exact same social activity. In column 7, we find a significant and negative effect. One reason might be because sports activities, unlike study activities, select individuals based on their physical proclivity, not cognitive capacities or educational preferences. Overall, we do not find that additional social interactions between neighbor-student pairs enhances the probability following one's older neighbors' college choices. We acknowledge our data limitations, but working with large administrative datasets for multiple cohorts, we shed some light on the effects of additional social interaction between neighbor-student pairs. To our knowledge, this is the first study that has provided an additional evidence of plausible peer-neighbor interactions in social groups, beyond the well-established evidence of neighbors living in close proximity.

6 Conclusion

We investigate individual neighbor spillover effects on specific college major choices. We find causal evidence that having an older neighbor who was admitted to and enrolled in a program increases a younger student's probability of applying to that program by about 20% (evaluated at the sample mean) compared to having a neighbor that barely missed the admission score. The effect is stronger if the student-neighbor pair lives in a city or region that does not have its own university, implying that the value of information in

rural areas is higher.

A heterogeneity analysis shows that female students are more likely to be influenced by older neighbors' choices, and male older neighbors' admission has a larger impact on both male and female students than that of female older neighbors. Furthermore, we find evidence that students don't follow their older neighbors to programs with low or medium demand, but have a higher probability of applying when neighbors are admitted to a high-demand (more prestige) program. Neighbor spillovers are smaller than sibling spillovers (as in Altmejd et al. 2021), which is also in line with findings by Avdeev et al. (2024) on field of study choice spillovers. Placebo tests confirm our findings. We reverse the order—i.e., we tested whether the college choices of younger neighbors affect the college choices of older neighbors—and find no significant results. Additionally, our findings are very robust to changes in specifications and sub-group analysis.

Next, we utilize the variation in study program weight schemes across Croatia's college system to identify the causal effect of older neighbors' admission to a college program on younger neighbors' human capital formation. We find a strong and positive effect through numerous robustness checks. After the older neighbor peer is admitted to a particular study program, younger neighbor peers tend to increase their grades in subjects needed to get into that college program relatively more than they do for the rest of their control-group subjects. The effect is stronger as the share of the weight placed on that particular subject in admissions increases. Moreover, we find that our results are not driven by the neighbor or younger students' specialization in a particular study field or a set of college programs with similar weight scheme. The driving effect occurs when older neighbors have a mixed college preference list (i.e., when a change in admission status for the older neighbor, from one ranked choice to the next, drastically changes the weight scheme). This serves as additional evidence that younger students follow the exact choices of older neighbors, not just within a general field of study. We show that younger neighbor peers learn the information about the specific subjects they need to excel in in order to increase their probability of admission to their older neighbor's college program. Whether this is a direct effect, via the older neighbors interacting with the younger ones, or an indirect effect, through aspiration, in which the older neighbor's admission motivates the younger neighbor to learn more about the admission process and consequently study more, is still an open question. These findings become even more relevant and applicable because, according to McGrath et al. 2014 and Haj et al. 2018, different types of hybrid college admission weight schemes are present in around 50 to 60 percent of college systems worldwide, across and within countries, and they are not solely an artifact of the Croatian college system.

Finally, we contribute to the literature by providing additional evidence of direct

exposure between neighbor-peers. We use a unique dataset on a universe of within- and outside-of-school activities for all students in Croatia for the observed period, which allows us to identify neighbor peer pairs who also participated in the same social activities. Due to rigorous and precise matching, our sample around the cutoff shrinks dramatically, and we find no enhancing effect of additional neighbor-peer interaction through participation in different social activities.

We show that information networks, which have been examined in previous studies, reach beyond just family networks. This is substantial from a policy perspective, as interventions to increase college enrollment and choices could have a multitude of effects beyond family boundaries. It is important to consider the potential channel of individual neighbor spillovers in educational trends, especially for places or regions where access to information is scarce. Neighborhoods matter, but in rural areas or regions where institutional settings are generally sparse, a few individuals could have a large impact on perceived and realized opportunities by those within their network when it comes to education and human capital formation.

References

- Abdulkadiroğlu, Atila, Joshua Angrist, and Parag Pathak (2014). “The elite illusion: Achievement effects at Boston and New York exam schools”. In: *Econometrica* 82.1, pp. 137–196.
- Altmejd, Adam, Andrés Barrios-Fernández, Marin Drlje, Joshua Goodman, Michael Hurwitz, Dejan Kovac, Christine Mulhern, Christopher Neilson, and Jonathan Smith (2021). “O brother, where start thou? Sibling spillovers on college and major choice in four countries”. In: *The Quarterly Journal of Economics* 136.3, pp. 1831–1886.
- Altonji, Joseph G, Erica Blom, and Costas Meghir (2012). “Heterogeneity in human capital investments: High school curriculum, college major, and careers”. In: *Annu. Rev. Econ.* 4.1, pp. 185–223.
- Altonji, Joseph G, Lisa B Kahn, and Jamin D Speer (2016). “Cashier or consultant? Entry labor market conditions, field of study, and career success”. In: *Journal of Labor Economics* 34.S1, S361–S401.
- Angrist, Joshua (2014). “The perils of peer effects”. In: *Labour Economics* 30, pp. 98–108.
- Angrist, Joshua and Alan Krueger (1991). “Does compulsory school attendance affect schooling and earnings?” In: *The Quarterly Journal of Economics* 106.4, pp. 979–1014.
- (1992). “The effect of age at school entry on educational attainment: an application of instrumental variables with moments from two samples”. In: *Journal of the American statistical Association* 87.418, pp. 328–336.
- (1994). “Why do World War II veterans earn more than nonveterans?” In: *Journal of labor economics* 12.1, pp. 74–97.
- (1995). “Split-sample instrumental variables estimates of the return to schooling”. In: *Journal of Business & Economic Statistics* 13.2, pp. 225–235.
- Ashenfelter, Orley and Joseph Mooney (1968). “Graduate education, ability, and earnings”. In: *The Review of Economics and Statistics*, pp. 78–86.
- Avdeev, Stanislav, Nadine Ketel, Hessel Oosterbeek, and Bas van der Klaauw (2024). “Spillovers in fields of study: Siblings, cousins, and neighbors”. In: *Journal of Public Economics* 238, p. 105193.
- Barrios-Fernández, Andrés (2022). “Neighbors’ Effects on University Enrollment”. In: *American Economic Journal: Applied Economics* 14.3, pp. 30–60.
- Breda, Thomas, Julien Grenet, Marion Monnet, and Clémentine Van Effenterre (2021). “Do female role models reduce the gender gap in science? Evidence from French high schools”. In.

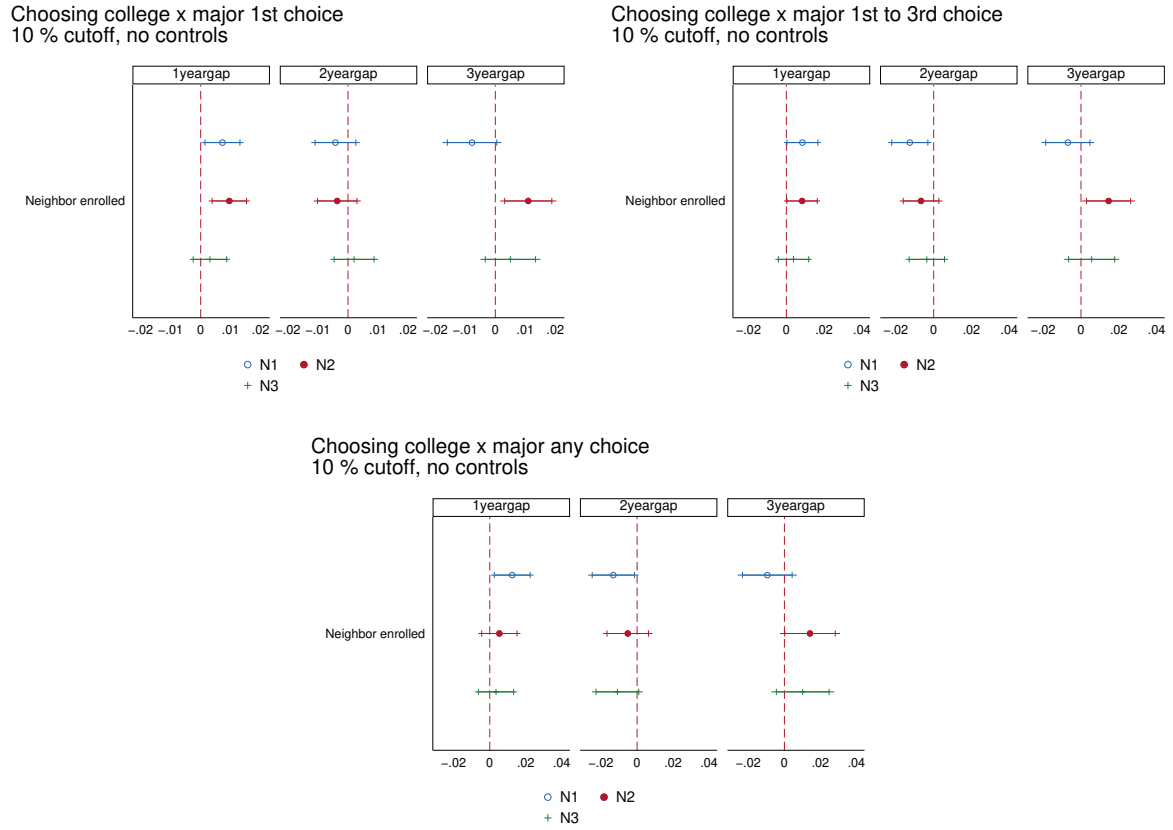
- Brooks-Gunn, Jeanne, Greg J Duncan, Pamela Kato Klebanov, and Naomi Sealand (1993). “Do neighborhoods influence child and adolescent development?” In: *American journal of sociology* 99.2, pp. 353–395.
- Calonico, Sebastian, Matias D Cattaneo, and Rocio Titiunik (2014). “Robust data-driven inference in the regression-discontinuity design”. In: *The Stata Journal* 14.4, pp. 909–946.
- Carlana, Michela, Eliana La Ferrara, and Paolo Pinotti (2022). “Goals and gaps: Educational careers of immigrant children”. In: *Econometrica* 90.1, pp. 1–29.
- Carrell, Scott and Bruce Sacerdote (2017). “Why do college-going interventions work?” In: *American Economic Journal: Applied Economics* 9.3, pp. 124–51.
- Chetty, Raj and Nathaniel Hendren (2018a). “The impacts of neighborhoods on intergenerational mobility I: Childhood exposure effects”. In: *The Quarterly Journal of Economics* 133.3, pp. 1107–1162.
- (2018b). “The impacts of neighborhoods on intergenerational mobility II: County-level estimates”. In: *The Quarterly Journal of Economics* 133.3, pp. 1163–1228.
- Chetty, Raj, Nathaniel Hendren, and Lawrence F Katz (2016). “The effects of exposure to better neighborhoods on children: New evidence from the moving to opportunity experiment”. In: *American Economic Review* 106.4, pp. 855–902.
- Chyn, Eric and Lawrence F Katz (2021). “Neighborhoods matter: Assessing the evidence for place effects”. In: *Journal of Economic Perspectives* 35.4, pp. 197–222.
- Ellen, Ingrid Gould and Margery Austin Turner (1997). “Does neighborhood matter? Assessing recent evidence”. In: *Housing policy debate* 8.4, pp. 833–866.
- Estrada, Ricardo, Jérémie Gignoux, and Agustina Hatrick (2021). “Learning About Opportunity: Spillovers of Elite School Admissions in Peru”. In: *CAF Working Paper*.
- Haj, Cezar Mihai, Irina Mihaela Geanta, and Dominic Orr (2018). “A typology of admission systems across Europe and their impact on the equity of access, progression and completion in higher education”. In: *European higher education area: The impact of past and future policies*, pp. 171–187.
- Hoxby, Caroline M and Christopher Avery (2013). “Low-income high-achieving students miss out on attending selective colleges”. In: *Brookings Papers on Economic Activity, Spring* 10.
- Jencks, Christopher, Susan E Mayer, et al. (1990). “The social consequences of growing up in a poor neighborhood”. In: *Inner-city poverty in the United States* 111, p. 186.
- Joensen, Juanna Schrøter and Helena Skyt Nielsen (2018). “Spillovers in education choice”. In: *Journal of Public Economics* 157, pp. 158–183.
- Kirkeboen, Lars J, Edwin Leuven, and Magne Mogstad (2016). “Field of study, earnings, and self-selection”. In: *The Quarterly Journal of Economics* 131.3, pp. 1057–1111.

- Manski, Charles F (1993). “Identification of endogenous social effects: The reflection problem”. In: *The review of economic studies* 60.3, pp. 531–542.
- (1995). *Identification problems in the social sciences*. Harvard University Press.
- McGrath, Cecile Hoareau, Marie Louise Henham, Anne Corbett, Niccolo Durazzi, Michael Frearson, Barbara Janta, Bregtje W Kamphuis, Eriko Katashiro, Nina Brankovic, Benoit Guerin, et al. (2014). “Higher education entrance qualifications and exams in Europe: A comparison”. In.
- Nakamura, Emi, Jósef Sigurdsson, and Jón Steinsson (2022). “The gift of moving: Inter-generational consequences of a mobility shock”. In: *The Review of Economic Studies* 89.3, pp. 1557–1592.
- Peter, Frauke H. and Vaishali Zambre (2017). “Intended college enrollment and educational inequality: Do students lack information?” In: *Economics of Education Review* 60, pp. 125–141.
- Qureshi, Javaeria A (2018). “Additional returns to investing in girls’ education: Impact on younger sibling human capital”. In: *The Economic Journal* 128.616, pp. 3285–3319.

A Appendix: Neighbor Proximity Effects

A.1 Main results for all neighbors (N=1 to N=3) and age difference (T=1 to T=3)

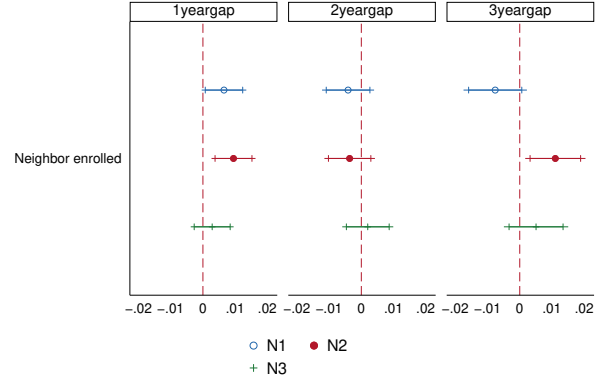
Figure 17: Results for all specifications, no control variables



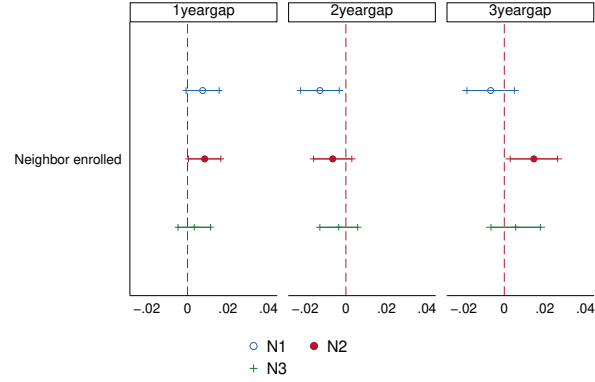
Notes: Displayed are, for each outcome of interest (applying as 1st preference, 1st to 3rd preference or any preference graphs in that order from top to bottom), coefficients of the IV estimations using the 10% bandwidth for the 3x3 matrix of all subsamples, for cohort differences of 1-3 years and first to third neighbor. All estimations include year and program fixed effects. Standard errors are clustered at student-pair level.

Figure 18: Results for all specifications

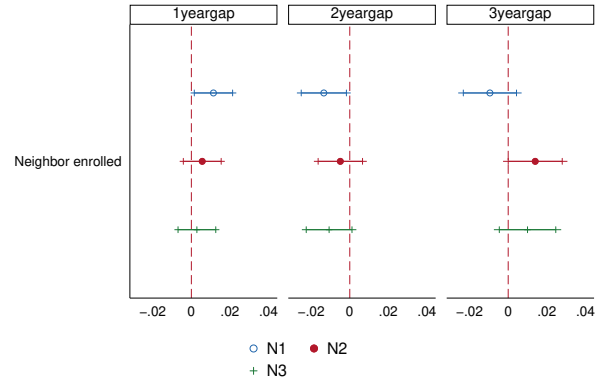
Choosing college x major 1st choice, 10 % cutoff



Choosing college x major 1st to 3rd choice, 10 % cutoff



Choosing college x major any choice, 10 % cutoff



Notes: Displayed are, for each outcome of interest (applying as 1st preference, 1st to 3rd preference or any preference graphs in that order from top to bottom), coefficients of the IV estimations using the 10% bandwidth for the 3x3 matrix of all subsamples, for cohort differences of 1-3 years and first to third neighbor. All estimations include year and program fixed effects and student characteristics as control variables. Standard errors are clustered at student-pair level.

Table 7: Placebo test: Second Neighbor and 1 Year Age Gap

	Neighbor admitted			Neighbor enrolled (fuzzy)		
	(1) Rdrobust	(2) OLS BW select	(3) OLS BW 10%	(4) Rdrobust	(5) IV BW select	(6) IV
Panel (a): Applying same program as 1st choice						
Robust	0.0112*** (0.004)			0.0113** (0.004)		
Neighbor admitted		0.0041 (0.004)	0.0048 (0.004)			
Neighbor enrolled					0.0044 (0.004)	0.0052 (0.004)
female = 1		-0.0096*** (0.002)	-0.0084*** (0.002)		-0.0095*** (0.002)	-0.0084*** (0.002)
same gender = 1		0.0134*** (0.002)	0.0132*** (0.002)		0.0143*** (0.002)	0.0131*** (0.002)
city has university		0.0050** (0.002)	0.0057** (0.002)		0.0055*** (0.002)	0.0056** (0.002)
distance in 1000m		0.0013** (0.001)	0.0010* (0.001)		0.0011* (0.001)	0.0010* (0.001)
Points_HS		0.0001*** (0.000)	0.0001*** (0.000)		0.0001*** (0.000)	0.0001*** (0.000)
Total_points		0.0000 (0.000)	0.0000 (0.000)		0.0000 (0.000)	0.0000 (0.000)
fstage					72670	60917
ymean	0.03	0.03	0.03	0.03	0.03	0.03
BW	71.44	71.44		80.88	80.88	
N	46023	28652	25879	46023	31037	25879
Panel (b): Applying same program as 1st - 3rd choice						
Robust	0.0098 (0.007)			0.0099 (0.007)		
Neighbor admitted		0.0036 (0.005)	0.0009 (0.005)			
Neighbor enrolled					0.0017 (0.006)	0.0010 (0.006)
female = 1		-0.0159*** (0.003)	-0.0138*** (0.003)		-0.0175*** (0.003)	-0.0138*** (0.003)
same gender = 1		0.0302*** (0.003)	0.0296*** (0.003)		0.0313*** (0.003)	0.0295*** (0.003)
city has university		0.0124*** (0.003)	0.0145*** (0.003)		0.0136*** (0.003)	0.0145*** (0.003)
distance in 1000m		0.0026*** (0.001)	0.0026*** (0.001)		0.0025*** (0.001)	0.0026*** (0.001)
Points_HS		0.0001*** (0.000)	0.0001*** (0.000)		0.0001*** (0.000)	0.0001*** (0.000)
Total_points		-0.0000*** (0.000)	-0.0000* (0.000)		-0.0000*** (0.000)	-0.0000* (0.000)
fstage					64311	60917
ymean	0.07	0.07	0.07	0.07	0.07	0.07
BW	64.24	64.24		69.85	69.85	
N	46023	26648	25879	46023	28237	25879
Panel (c): Applying same program any choice						
Robust	0.0082 (0.009)			0.0083 (0.009)		
Neighbor admitted		0.0026 (0.007)	-0.0029 (0.007)			
Neighbor enrolled					-0.0010 (0.007)	-0.0031 (0.007)
female = 1		-0.0208*** (0.004)	-0.0194*** (0.004)		-0.0227*** (0.004)	-0.0194*** (0.004)
same gender = 1		0.0506*** (0.004)	0.0496*** (0.004)		0.0516*** (0.004)	0.0496*** (0.004)
city has university		0.0121*** (0.004)	0.0135*** (0.004)		0.0115*** (0.004)	0.0135*** (0.004)
distance in 1000m		0.0029** (0.001)	0.0029** (0.001)		0.0029** (0.001)	0.0029** (0.001)
Points_HS		0.0001*** (0.000)	0.0001*** (0.000)		0.0001*** (0.000)	0.0001*** (0.000)
Total_points		-0.0000 (0.000)	-0.0000 (0.000)		-0.0000 (0.000)	-0.0000 (0.000)
BW	59.87	59.87		66.54	66.54	
fstage					61453	60917
ymean	0.12	0.11	0.11	0.12	0.11	0.11
N	46023	25283	25879	46023	27367	25879

Notes: Placebo test: the effect of a younger student's admission and enrollment to a program on older neighbor's choice set. Columns (1)-(3) show the reduced form results for the student's admission, columns (4)-(6) for the student's enrollment. Panels (a)-(c) show results for dependent variables choosing the same program as 1st preference, as 1st-3rd preference or any preference. Columns 1 and 4 show results for the rdrobust estimation and the resulting bandwidths. Column 2 and 5 show results of the linear regression using the data driven bandwidths from the preceding model, Columns 3 and 6 the results when using 10% bandwidths relative to the total admission score per program. All regressions include college \times major and year fixed effects. Standard errors are clustered at student-neighbor pair level.* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table 8: Placebo test: Third Neighbor and 1 Year Age Gap

	Neighbor admitted			Neighbor enrolled (fuzzy)		
	(1) Rdrobust	(2) OLS BW select	(3) OLS BW 10%	(4) Rdrobust	(5) IV BW select	(6) IV
Panel (a): Applying same program as 1st choice						
Robust	0.0098* (0.005)			0.0103* (0.005)		
Neighbor admitted		0.0047 (0.004)	0.0055 (0.004)			
Neighbor enrolled				0.0053 (0.004)	0.0060 (0.005)	
female = 1		-0.0116*** (0.003)	-0.0100*** (0.003)	-0.0118*** (0.002)	-0.0099*** (0.003)	
same gender = 1		0.0175*** (0.002)	0.0158*** (0.003)	0.0177*** (0.002)	0.0158*** (0.003)	
city has university		0.0052** (0.002)	0.0054** (0.003)	0.0054** (0.002)	0.0054** (0.003)	
distance in 1000m		0.0011 (0.001)	0.0015* (0.001)	0.0012* (0.001)	0.0014* (0.001)	
Points_HS		0.0001*** (0.000)	0.0001*** (0.000)	0.0001*** (0.000)	0.0001*** (0.000)	
Total_points		0.0000 (0.000)	0.0000* (0.000)	0.0000 (0.000)	0.0000* (0.000)	
fstage					47730	39643
ymean	0.03	0.03	0.03	0.03	0.03	0.03
BW	77.88	77.88		82.80	82.80	
N	30249	20052	17330	30249	20755	17330
Panel (b): Applying same program as 1st - 3rd choice						
Robust	0.0086 (0.008)			0.0099 (0.009)		
Neighbor admitted		0.0001 (0.006)	0.0018 (0.007)			
Neighbor enrolled				0.0021 (0.007)	0.0020 (0.007)	
female = 1		-0.0217*** (0.004)	-0.0185*** (0.004)	-0.0223*** (0.004)	-0.0185*** (0.004)	
same gender = 1		0.0329*** (0.004)	0.0294*** (0.004)	0.0327*** (0.004)	0.0294*** (0.004)	
city has university		0.0086** (0.004)	0.0073* (0.004)	0.0091** (0.004)	0.0073* (0.004)	
distance in 1000m		0.0031*** (0.001)	0.0035*** (0.001)	0.0030*** (0.001)	0.0035*** (0.001)	
Points_HS		0.0001*** (0.000)	0.0001*** (0.000)	0.0001*** (0.000)	0.0001*** (0.000)	
Total_points		0.0000 (0.000)	0.0000 (0.000)	0.0000 (0.000)	0.0000 (0.000)	
fstage					40936	39643
ymean	0.07	0.07	0.06	0.07	0.07	0.06
BW	77.22	77.22		69.83	69.83	
N	30249	19738	17330	30249	18669	17330
Panel (c): Applying same program any choice						
Robust	0.0135 (0.009)			0.0157 (0.011)		
Neighbor admitted		0.0037 (0.007)	0.0035 (0.008)			
Neighbor enrolled				0.0053 (0.009)	0.0038 (0.009)	
female = 1		-0.0287*** (0.005)	-0.0250*** (0.005)	-0.0292*** (0.005)	-0.0250*** (0.005)	
same gender = 1		0.0564*** (0.005)	0.0508*** (0.005)	0.0540*** (0.005)	0.0508*** (0.005)	
city has university		0.0094* (0.005)	0.0065 (0.005)	0.0085 (0.005)	0.0065 (0.005)	
distance in 1000m		0.0033** (0.001)	0.0035** (0.001)	0.0026* (0.001)	0.0035** (0.001)	
Points_HS		0.0001*** (0.000)	0.0001*** (0.000)	0.0001*** (0.000)	0.0001*** (0.000)	
Total_points		0.0000 (0.000)	-0.0000 (0.000)	-0.0000 (0.000)	-0.0000 (0.000)	
BW	81.34	81.34		69.20	69.20	
fstage					40630	39643
ymean	0.12	0.11	0.11	0.12	0.11	0.11
N	30249	20379	17330	30249	18573	17330

Notes: Placebo test: the effect of a younger student's admission and enrollment to a program on older neighbor's choice set. Columns (1)-(3) show the reduced form results for the student's admission, columns (4)-(6) for the student's enrollment. Panels (a)-(c) show results for dependent variables choosing the same program as 1st preference, as 1st-3rd preference or any preference. Columns 1 and 4 show results for the rdrobust estimation and the resulting bandwidths. Column 2 and 5 show results of the linear regression using the data driven bandwidths from the preceding model, Columns 3 and 6 the results when using 10% bandwidths relative to the total admission score per program. All regressions include college \times major and year fixed effects. Standard errors are clustered at student-neighbor pair level. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

B College Program Weight Scheme in Croatia

B.1 Weight of State Exams

We utilize the weight scheme from the centralized college system for 2022. According to Agency for Science and Higher Education data, the weight scheme must go through internal review at the university level for every college program, and then be validated by the Ministry of Science. The changes are usually marginal and not structural. For instance, a physics study program may shift the balance of physics and mathematics weights, but it will not change its weights toward social science subjects such as history or geography.

B.2 Subjects and Weights Omitted

When Croatian high school students take state exams, they can choose (to some extent) how many subjects, and in which subjects they take the state exam in. The exception is math and Croatian, which are mandatory, so students can only decide between A- and B-level courses and exams. Therefore, we exclude math and Croatian from the list of subjects included in the analysis for two reasons:

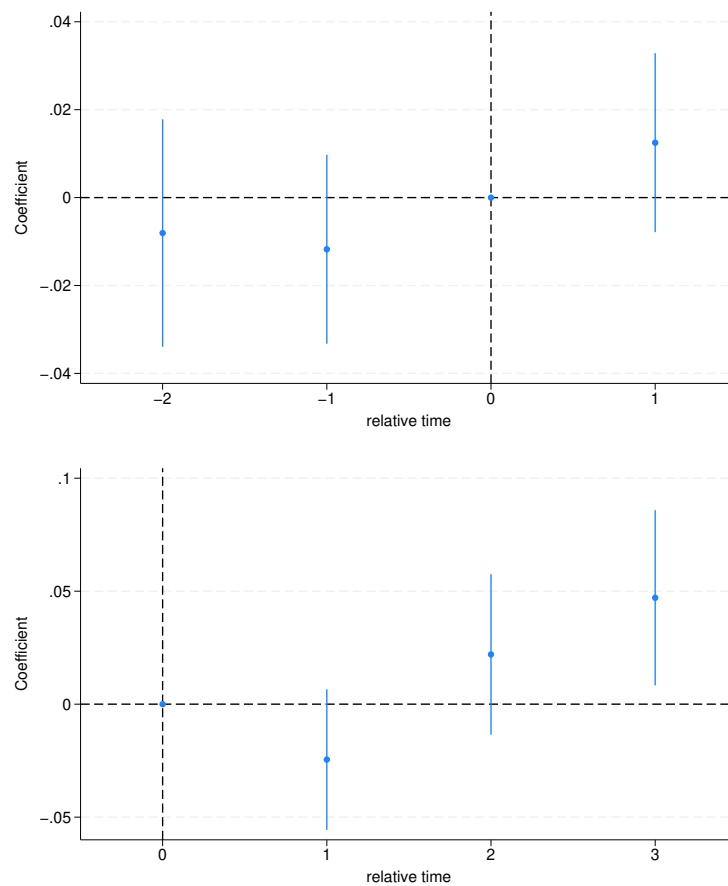
- For most study programs, a mandatory condition is at least a passing grade on math and Croatian state exams, and for a large share of programs, school grades in these subjects also hold weight in the calculation of the application score. Observations for these subjects would therefore be included in the treatment group for nearly all students.
- As all students take a state exam in these subjects, it's likely that their school grades might improve anyway, through studying, as they get closer to the state exam. Improvement in grades would therefore not be an effect coming from an older neighbor.

The treatment group subjects are, after excluding math and Croatian, defined by the different weights of state exam subjects. Grades of all subjects that are required to be at least passed in the state exam therefore make up the control group. We only include observations of students where the program the older neighbor got admitted to requires at least one subject other than math or Croatian. The control group consists of grades of other subjects. Here, we choose only subjects that are also an important part of general education and also "typical" state exam subjects, such as languages, chemistry, geography, etc., if not already included in the subject treatment group. We exclude arts subjects and physical education, as well as highly specialized subjects only offered at some schools.

We then run a simple difference-in-difference specification using the within-student variation. We include student fixed effects and cluster at student \times treatment level. We control for calendar year and subject-specific differences. We only include subjects per student where we have grade information for all class levels and exclude students that repeated a grade for this analysis.

B.3 Event Study DiD T1 and T3

Figure 19: Differences in grades for subjects both important and unimportant for program older neighbor got admitted to



Notes: DiD coefficients for the closest neighbor ($N=1$) with one and three years difference in age. $T=0$ is the school year when older neighbor was admitted to a particular college program. College admissions are in July, while the school year ends in June, so this event happens a month after younger student finished 9th and 11th grade.

Table 9: Corresponding regression results N1T1

	(1) main b/se	(2) Q1 weights b/se	(3) Q4 weights b/se	(4) Q1 GPA b/se	(5) Q4 GPA b/se
0.ttt	0.0160* (0.009)	-0.0037 (0.018)	0.0662*** (0.019)	0.0427** (0.018)	-0.0095 (0.008)
1.ttt	-0.0084 (0.006)	-0.0152 (0.012)	0.0149 (0.015)	0.0154 (0.014)	-0.0116* (0.007)
3.ttt	0.1247*** (0.006)	0.1245*** (0.011)	0.1473*** (0.014)	0.0885*** (0.013)	0.0523*** (0.006)
1.subjT	0.0049 (0.012)	0.0186 (0.023)	0.0142 (0.030)	-0.0041 (0.023)	0.0355** (0.015)
0.ttt \times 1.subjT	-0.0081 (0.013)	-0.0160 (0.025)	-0.0378 (0.030)	-0.0215 (0.024)	-0.0071 (0.016)
1.ttt \times 1.subjT	-0.0118 (0.011)	0.0013 (0.021)	-0.0391 (0.026)	0.0018 (0.021)	-0.0156 (0.015)
3.ttt \times 1.subjT	0.0125 (0.010)	0.0073 (0.020)	-0.0137 (0.025)	-0.0110 (0.022)	-0.0040 (0.014)
_cons	4.2397*** (0.022)	4.2567*** (0.041)	4.3124*** (0.054)	3.4865*** (0.056)	4.8228*** (0.019)
<i>N</i>	121752	37196	22168	30343	33863

Notes: All regressions include Student fixed effect, are clustered at Student \times subjT level, and include year and subject controls. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table 10: Corresponding regression results N1T3

	(1) main b/se	(2) Q1 weights b/se	(3) Q4 weights b/se	(4) Q1 GPA b/se	(5) Q4 GPA b/se
1.ttt	-0.0194** (0.010)	-0.0083 (0.021)	-0.0242 (0.016)	-0.0141 (0.017)	0.0044 (0.009)
2.ttt	-0.0274** (0.012)	0.0115 (0.029)	-0.0321 (0.021)	-0.0235 (0.021)	0.0154 (0.011)
3.ttt	0.0885*** (0.014)	0.1680*** (0.034)	0.0890*** (0.025)	0.0458* (0.024)	0.0528*** (0.012)
1.subjT	0.0157 (0.018)	0.0848** (0.037)	-0.0232 (0.037)	0.0182 (0.028)	0.0197 (0.022)
1.ttt×1.subjT	-0.0245 (0.016)	-0.0828** (0.033)	0.0409 (0.030)	-0.0215 (0.026)	-0.0052 (0.023)
2.ttt×1.subjT	0.0220 (0.018)	-0.0805** (0.040)	0.0928*** (0.033)	0.0210 (0.027)	0.0124 (0.024)
3.ttt×1.subjT	0.0471** (0.020)	-0.0275 (0.044)	0.0764** (0.036)	0.0135 (0.030)	0.0295 (0.024)
_cons	4.2705*** (0.029)	4.0921*** (0.056)	4.3295*** (0.056)	3.7161*** (0.055)	4.8522*** (0.021)
<i>N</i>	62176	14560	18740	24978	16357

Notes: All regressions include Student fixed effect, are clustered at Student \times subjT level, and include year and subject controls. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table 11: DiD results N1T1

	(1) main b/se	(2) Q1 weights b/se	(3) Q4 weights b/se	(4) Q1 GPA b/se	(5) Q4 GPA b/se
1.timeT	0.1355*** (0.007)	0.1236*** (0.013)	0.1779*** (0.017)	0.1064*** (0.016)	0.0475*** (0.007)
1.subjT	-0.0017 (0.010)	0.0137 (0.020)	-0.0114 (0.026)	-0.0101 (0.019)	0.0281** (0.011)
1.timeT×1.subjT	0.0192* (0.011)	0.0122 (0.021)	0.0120 (0.025)	-0.0049 (0.020)	0.0034 (0.012)
_cons	4.2727*** (0.026)	4.2439*** (0.056)	4.4459*** (0.065)	3.5663*** (0.061)	4.8014*** (0.024)
<i>N</i>	121752	37196	22168	30343	33863

Notes: full sample for T1 is comparison for grades of subjects mandatory/ with weights of the program the older student got admitted to vs. grades of other subjects, within-student variation, Q1 means students were weights make up only 40 percent of the total final score, Q4 60 or more percent of final score. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

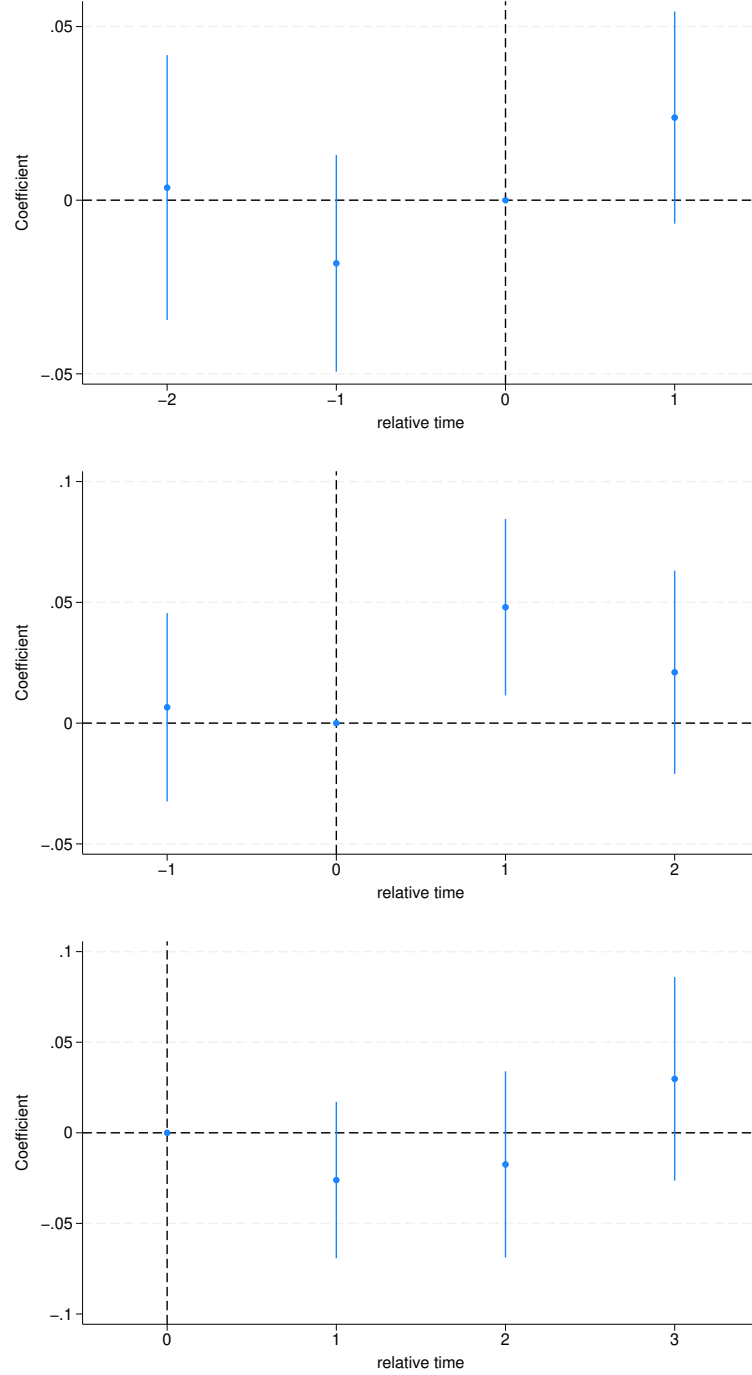
Table 12: DiD results N1T3

	(1)	(2)	(3)	(4)	(5)
	main	Q1 weights	Q4 weights	Q1 GPA	Q4 GPA
	b/se	b/se	b/se	b/se	b/se
1.timeT	-0.1018*** (0.012)	-0.1211*** (0.029)	-0.0988*** (0.022)	-0.0631*** (0.020)	-0.0304*** (0.012)
1.subjT	0.0155 (0.018)	0.0836** (0.037)	-0.0230 (0.037)	0.0132 (0.027)	0.0198 (0.022)
1.timeT \times 1.subjT	0.0151 (0.015)	-0.0620* (0.034)	0.0697** (0.029)	0.0066 (0.023)	0.0134 (0.020)
_cons	4.1961*** (0.032)	3.9598*** (0.071)	4.2577*** (0.062)	3.7155*** (0.058)	4.8164*** (0.025)
<i>N</i>	62176	14560	18740	26422	16357

Notes: full sample for T3 is comparison for grades of subjects mandatory/ with weights of the program the older student got admitted to vs. grades of other subjects, within-student variation, Q1 means students were weights make up only 40 percent of the total final score, Q4 60 or more percent of final score. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

C DiD by Type of Student and Neighbor Preference List

Figure 20: Further Placebo: N1 T1-T3 of not admitted older neighbors



Notes: DiD coefficients for the closest neighbor (N=1) with one, two, and three years difference in age for not-admitted older neighbors. T=0 is the school year when older neighbor was admitted to a particular college program. College admissions are in July, while the school year ends in June, so this event happens a month after younger student finished 9th,10th or 11th grade.

Table 13: Effect N1T2 by student preference strength

	(1)	(2)	(3)	(4)	(5)	(6)
	citySq1	citySq4	fieldSq1	fieldSq4	uniSq1	uniSq4
	b/se	b/se	b/se	b/se	b/se	b/se
1.timeT	-0.0459**	-0.0705***	-0.0726***	-0.0352	-0.0978***	-0.0615***
	(0.018)	(0.017)	(0.020)	(0.022)	(0.021)	(0.016)
1.subjT	-0.0442	0.0312	-0.0412	-0.0238	-0.0341	-0.0019
	(0.027)	(0.024)	(0.030)	(0.030)	(0.031)	(0.022)
1.timeT \times 1.subjT	0.0260	0.0553**	0.0597**	0.0426	0.0228	0.0611***
	(0.025)	(0.023)	(0.028)	(0.028)	(0.028)	(0.021)
_cons	4.0304***	4.1662***	4.1185***	4.2475***	4.1116***	4.0987***
	(0.060)	(0.056)	(0.064)	(0.067)	(0.065)	(0.050)
<i>N</i>	21224	25012	16168	16504	15912	29692

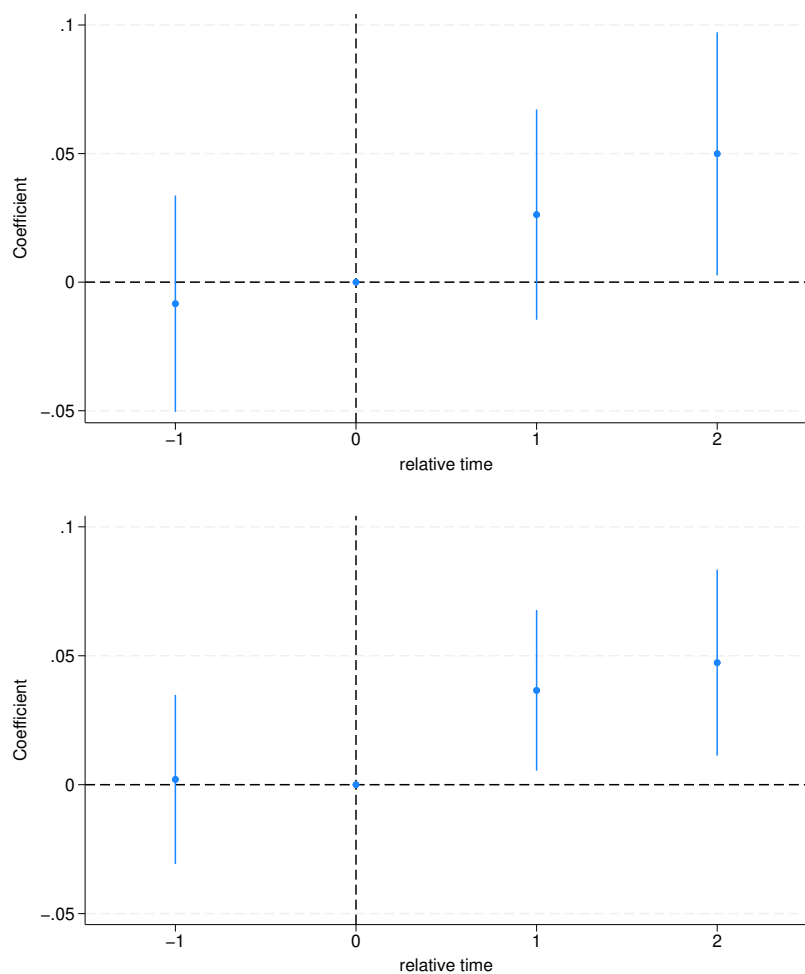
Notes: Preference identification of student. Low value (q1) means that the student has a very mixed application choice list. Values close to 1 represent a very concentrated choice set. We look at the choice set mixed vs not in regard to variety in city choices, field of study choices, and university choices. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table 14: Effect N1T2 by neighbor preference strength

	(1)	(2)	(3)	(4)	(5)	(6)
	cityNq1	cityNq4	fieldNq1	fieldNq4	uniNq1	uniNq4
	b/se	b/se	b/se	b/se	b/se	b/se
1.timeT	-0.0600***	-0.0324*	-0.0729***	-0.0436**	-0.0621***	-0.0235
	(0.020)	(0.017)	(0.020)	(0.020)	(0.020)	(0.016)
1.subjT	-0.0789***	0.0166	0.0796***	-0.0816***	-0.0876***	0.0352
	(0.030)	(0.024)	(0.029)	(0.030)	(0.029)	(0.023)
1.timeT \times 1.subjT	0.0128	0.0462**	0.0167	0.0279	0.0328	0.0476**
	(0.028)	(0.023)	(0.029)	(0.026)	(0.026)	(0.023)
_cons	4.1604***	3.9853***	4.1187***	4.0482***	4.1553***	4.0030***
	(0.057)	(0.058)	(0.062)	(0.063)	(0.051)	(0.068)
<i>N</i>	18044	23880	17392	19220	19156	25180

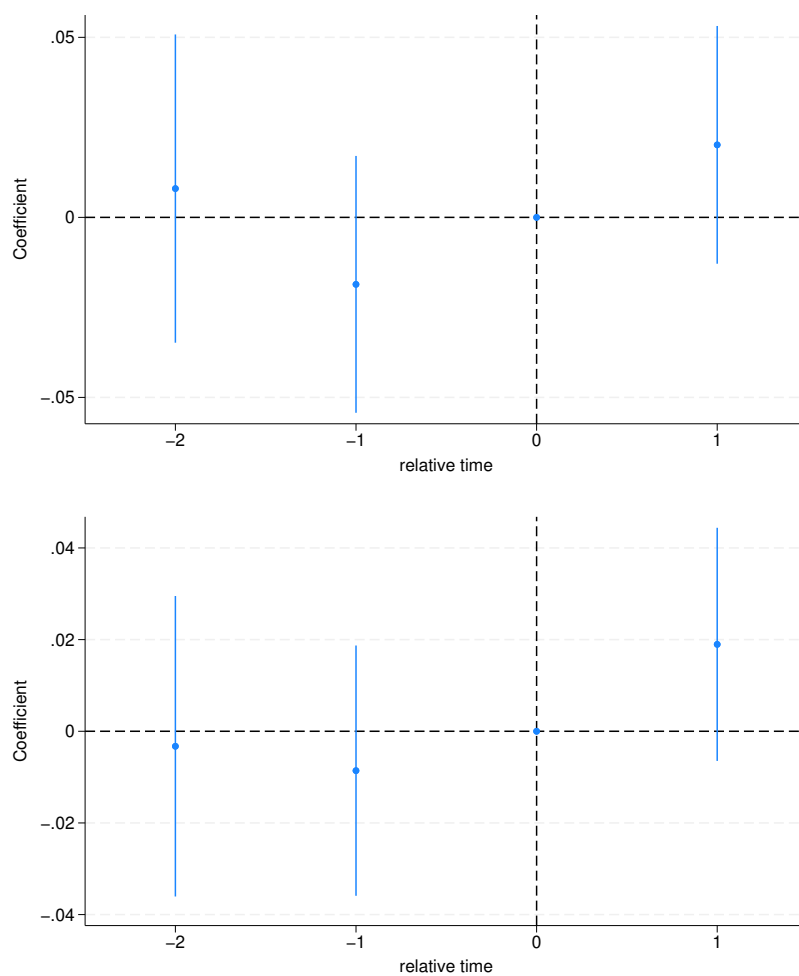
Notes: Preference identification of neighbor. Low value (q1) means that the student has a very mixed application choice list. Values close to 1 represent a very concentrated choice set. We look at the choice set mixed vs not in regard to variety in city choices, field of study choices, and university choices. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Figure 21: N1T2 only admitted neighbors and only 5% and 10% above cutoff (in bandwidth)



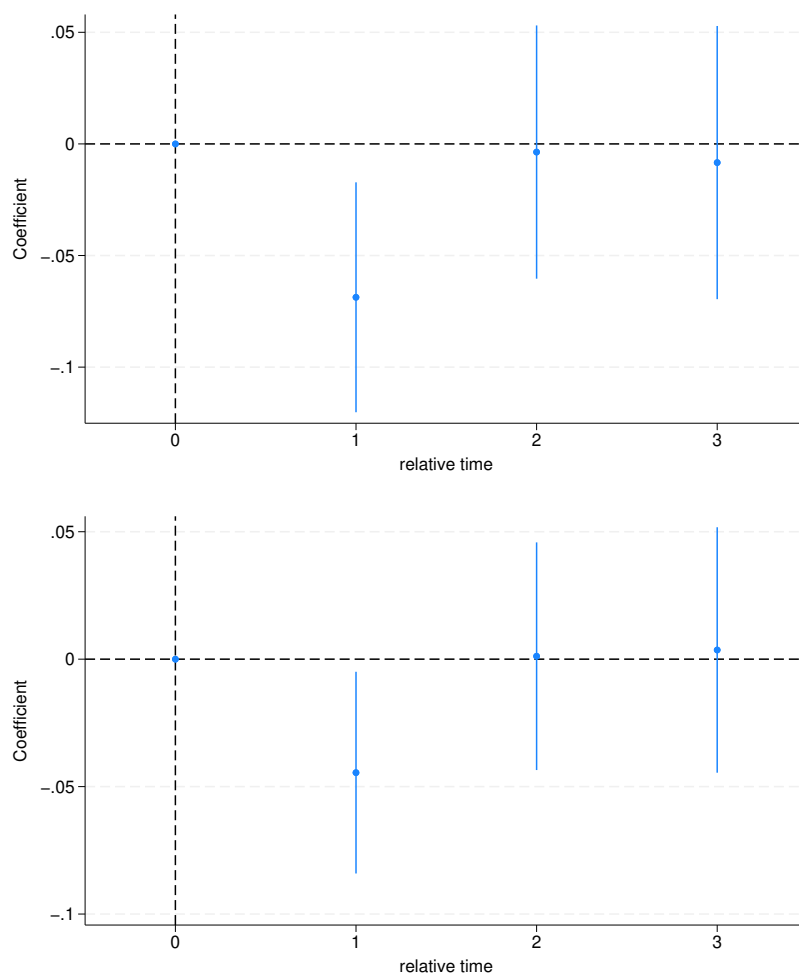
Notes: DiD coefficients for the closest neighbor ($N=1$) with two years difference in age. $T=0$ is the school year when older neighbor was admitted to a particular college program. College admissions are in July, while the school year ends in June, so this event happens a month after younger student finished 10th grade.

Figure 22: N1T1 only admitted neighbors and only 5% and 10% above cutoff (in bandwidth)



Notes: DiD coefficients for the closest neighbor (N=1) with one year difference in age. T=0 is the school year when older neighbor was admitted to a particular college program. College admissions are in July, while the school year ends in June, so this event happens a month after younger student finished 11th grade.

Figure 23: N1T3 only admitted neighbors and only 5% and 10% above cutoff (in bandwidth



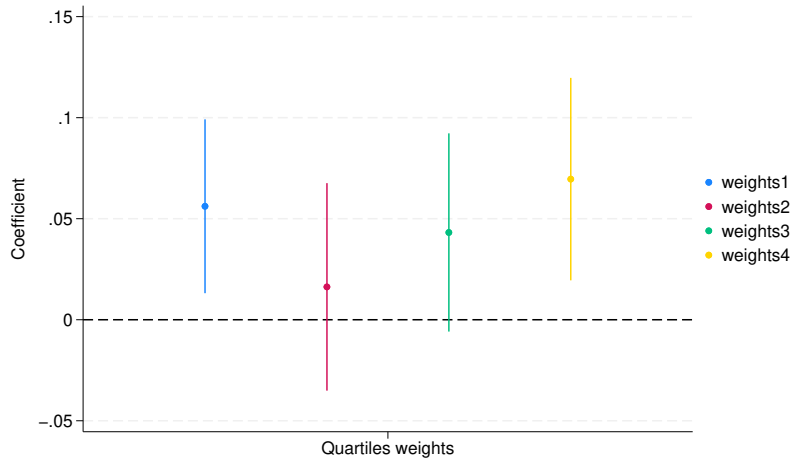
Notes: DiD coefficients for the closest neighbor ($N=1$) with three years difference in age. $T=0$ is the school year when older neighbor was admitted to a particular college program. College admissions are in July, while the school year ends in June, so this event happens a month after younger student finished 9th grade.

Table 15: DiD Results for N=1 and T=2 (alternative measure)

	(1)	(2)	(3)	(4)	(5)
	main	Q1 weights	Q4 weights	Q1 GPA	Q4 GPA
	b/se	b/se	b/se	b/se	b/se
1.timeT	-0.0534*** (0.009)	-0.0267 (0.017)	-0.0686*** (0.018)	-0.0687*** (0.020)	-0.0158 (0.010)
1.subjT	-0.0168 (0.013)	-0.0177 (0.022)	-0.0322 (0.028)	-0.0301 (0.024)	-0.0131 (0.015)
1.timeT×1.subjT	0.0485*** (0.012)	0.0562** (0.022)	0.0647*** (0.023)	0.0458** (0.021)	0.0377*** (0.014)
_cons	4.1124*** (0.030)	4.0809*** (0.057)	4.2080*** (0.047)	3.3584*** (0.065)	4.8148*** (0.026)
N	83404	27388	22956	20777	23617

Notes: Full sample for T2 is comparison for grades of subjects mandatory/ with weights of the program the older student got admitted to vs grades of other subjects, within student variation, Q1 means students where subject weights of the neighbor's program are zero (but subjects are mandatory to pass, which is the case for about 25% of students in the sample), Q4 the weights of the subjects make up 40% or more of the final grade. For the previous table, the overall weight of all subjects was used. In this case we only look at the weights of the subjects that were included in the treatment group in the end. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Figure 24: DiD Coefficients by quartiles of program weight dispersion



Notes: This shows the results by program weight quartiles corresponding to the alternative main Table 15. Quartiles are distributed according to the weights of the treatment subjects.

Halle Institute for Economic Research –
Member of the Leibniz Association

Kleine Maerkerstrasse 8
D-06108 Halle (Saale), Germany

Postal Address: P.O. Box 11 03 61
D-06017 Halle (Saale), Germany

Tel +49 345 7753 60
Fax +49 345 7753 820

www.iwh-halle.de

ISSN 2194-2188



The IWH is funded by the federal government and the German federal states.