CESifo WORKING PAPERS

9181
2021

July 2021

Age at Arrival and Residential Integration

Cristina Bratu, Matz Dahlberg, Madhinee Valeyatheepillay

CESifo

Impressum:

CESifo Working Papers
ISSN 2364-1428 (electronic version)
Publisher and distributor: Munich Society for the Promotion of Economic Research - CESifo GmbH
The international platform of Ludwigs-Maximilians University's Center for Economic Studies and the ifo Institute
Poschingerstr. 5, 81679 Munich, Germany
Telephone +49 (0)89 2180-2740, Telefax +49 (0)89 2180-17845, email office@cesifo.de
Editor: Clemens Fuest
https://www.cesifo.org/en/wp
An electronic version of the paper may be downloaded
· from the SSRN website: www.SSRN.com
· from the RePEc website: www.RePEc.org
· from the CESifo website: https://www.cesifo.org/en/wp

CESifo *Working Paper No. 9181*

Age at Arrival and Residential Integration

Abstract

We study residential integration patterns in adulthood for children of refugees who arrive in Sweden before the age of 16. Using geo-coded information on the residential location of each individual in Sweden, we take a novel, data-driven approach in defining neighborhoods and construct individualized k-nearest neighborhoods, for k = 100 or k = 1000. Exploiting a siblings design, we find that, at age 30, refugee children arriving later live in neighborhoods with lower shares of natives, high-educated individuals, and high-income earners, and higher share of welfare receivers, regardless of the level of k. We also provide evidence that refugee children arriving later experience worse labor market outcomes in terms of earnings, lower educational outcomes and likelihood to marry Swedish-born partners at age 30 as compared to children arriving earlier to the host country. Using a decomposition analysis, we show that the mean effects of age at arrival on neighborhood integration are only partly explained by economic integration, educational integration and intermarriage. Our findings indicate that a large part of the estimated mean age at arrival effects remains unaccounted for, particularly for k = 100, which suggests a role for Swedish housing policies, housing discrimination and taste-based preferences in fully explaining the effects of age at arrival.

JEL-Codes: R230, J150, J120, J010.

Keywords: refugees, residential integration, age at arrival.

*Cristina Bratu**
Aalto University School of Business
Aalto / Finland
cristina.bratu@aalto.fi

Matz Dahlberg
Uppsala University
Uppsala / Sweden
matz.dahlberg@ibf.uu.se

Madhinee Valeyatheepillay
ifo Institute – Leibniz Institute for Economic Research
at the University of Munich / Germany
valeyatheepillay@ifo.de

*corresponding author

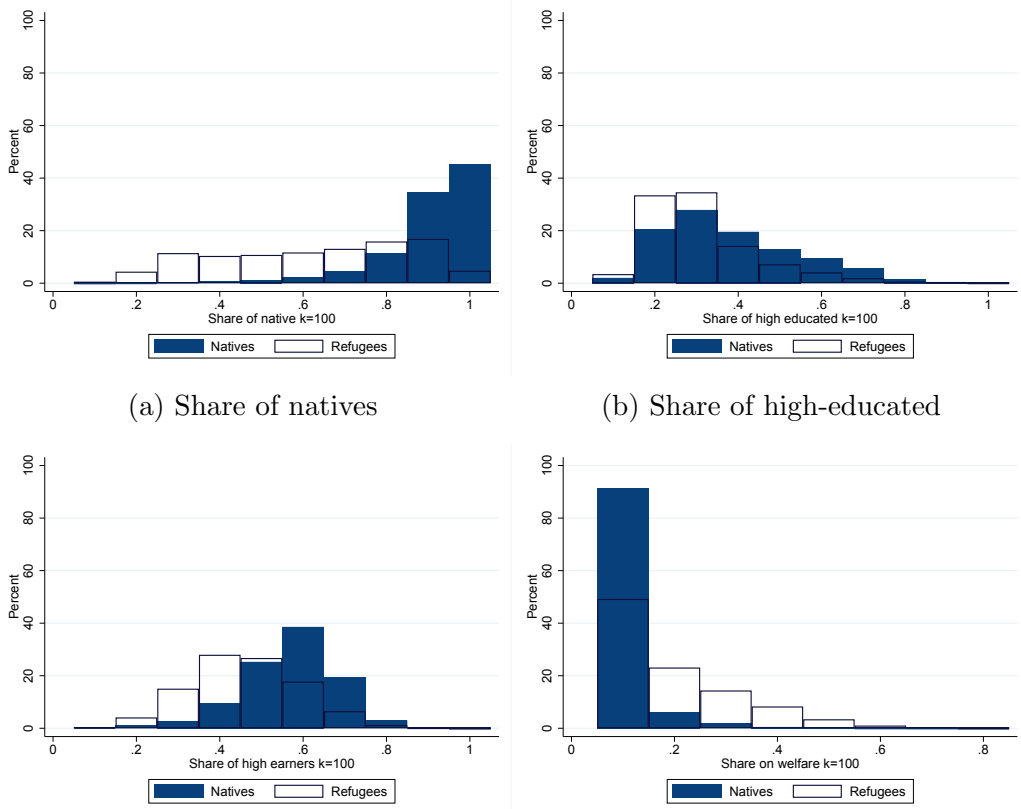
July 5, 2021

1 Introduction

Refugees live in vastly different neighborhoods compared to natives. Figure 1 shows, for example, that while the majority of natives in Sweden live in neighborhoods where 90% of their 100 closest neighbors are natives, only around 20% of refugees live in similarly native-dominant neighborhoods. The purpose of this paper is to examine whether being exposed earlier and for longer to the host country could explain the large variation observed in the type of neighborhoods refugees live in.

We hypothesize that the younger refugees are upon arrival, the more time they have to build country-specific knowledge, including language and culture, and to forge social contacts with the native majority, which may affect both their preferences for certain kinds of neighborhoods and their ability to act upon those preferences.

Figure 1: Characteristics of 100 closest neighbors for natives and refugees in 2014



(a) Share of natives

(b) Share of high-educated

(c) Share of high-earners

(d) Share on welfare

Notes: The figure shows the characteristics of the 100 closest neighbors for all refugees and natives above the age of 18 who were residing in Sweden in 2014. Refugees are defined based on residence permit data. Natives are individuals born in Sweden; high-educated individuals have at least some tertiary education; high-earners are defined as earning above the median in the municipality; on welfare refers to receipt of social benefits.
Source: Own calculations on data from the GeoSweden database.

We use administrative data to study refugees born between 1974 and 1984 who arrive in Sweden before the age of 16 and whose residential locations can be observed at age 30. Using geo-coded information on the residential location of each individual in Sweden, - given by 100 $\times$ 100m coordinates - we construct individualized k-nearest neighborhoods, for values of k equal to 100 and 1000. This method allows us to identify the characteristics of neighbors at both very granular levels and at more aggregate levels. We analyze the extent to which age at immigration affects neighborhood composition along two dimensions: i) ethnic composition, measured as the share of natives, defined as individuals born in Sweden, and ii) socio-economic composition, measured via three variables: the share of high-income earners, the share of high-educated individuals, and the share of individuals on welfare.[1] We apply a siblings design to estimate the effect of arriving at different ages relative to a reference group that arrives between the ages of 0 and 3. The within-family analysis enables us to address potential selection bias stemming from the fact that parents with better unobservables may move abroad when their children are younger.[2] We provide suggestive evidence for the mechanisms that generate these outcomes by performing a decomposition analysis in the style of Heckman et al. (2013) to analyse how much of the effect of age at arrival on neighborhood integration goes through earnings, education and intermarriage, which is defined as being married to or cohabiting (with children) with a Swedish-born partner.

Our baseline results show that compared to refugee children arriving between the ages of 0 and 3, refugee children arriving later experience a larger deviation from natives in terms of the composition of their neighbors at age 30. The effects on residential integration both along the ethnic and socio-economic lines are flat until around school-starting age, when they start declining. There are no marked differences between $k = 100$ and $k = 1000$. The effects are large. For example, those that arrive at age 15 live in neighborhoods with a 7 percentage points lower share of natives among their closest neighbors, which amounts to 35 percent of the mean value for the reference group. The corresponding magnitudes for the socio-economic characteristics of their neighbors are approximately 6 percentage points (share high-educated), 7.5 percentage points (share of high-income earners), and 7.5 percentage points (share on welfare).

We next show that age at arrival negatively affects refugees' labor market integration – as measured by income rank and years of education – and the probability

[1] We define these variables more precisely in Section 2.

[2] We note, however, that such issues are likely to be less prevalent in our sample of refugees, who are more likely to move so as to escape violence and conflict, and thus have less control over the timing of their moves.

of marrying a native. The estimated effects are sizeable. For instance, arriving in Sweden at age 15 rather than at ages 0-3 leads to approximately a 12.5 lower percentile rank in the earnings distribution at age 30, a half a year less of education, and a 28 percentage point lower probability of being married to a native-born partner (conditional on being married).[3]

Finally, we decompose the baseline results in order to assess how much of the effects of age at arrival on residential integration operate through the labor market and education channels and how much through the intermarriage channel. We find that income rank, years of education and intermarriage contribute between about 20 to 40 percent of the variation in neighborhood characteristics. However, a large part of the effects of age at arrival on residential outcomes remains unexplained, particularly for very small neighborhoods ($k = 100$).

Previous literature shows that neighborhoods matter for several reasons. First, neighbors can have a direct impact on their neighborhood peers' key life outcomes, such as labor market and educational outcomes, by sharing information, resources and knowledge, and by influencing various types of behaviors and attitudes, such as voting behavior (see, e.g., Borjas, 1995, Ellen and Turner, 1997, List et al., 2020, Sampson et al., 2002, Johnston and Pattie, 2011, Sharkey and Faber, 2014, Graham, 2018, Chetty et al., 2020). Moreover, childhood environments shape children's long-run outcomes: children who grow up in poor neighborhoods experience worse labor market outcomes in adulthood (Chetty et al., 2016). Second, intergroup contact at the neighborhood-level increases trust (Dinesen and Sønderskov, 2018), and trust facilitates integration (Nannestad, 2004, 2009). Third, there may be large disparities in available resources and services in neighborhoods in which different groups live (De la Roca et al., 2018). Finally, having non-segregated neighborhoods remains an important political goal in many countries.

We make several contributions. To our knowledge, we are the first to examine a determinant of small-scale neighborhood integration. Our flexible neighborhood definition is based on a k-nearest neighbor approach. This approach presents several advantages: we can create neighborhoods with constant counts of individuals as compared to administrative units. Furthermore, our approach can better capture what refugees identify as their neighborhood, because it puts the refugee at the center of their own neighborhood. Most importantly, we can conduct small scale neighborhood analysis, down to $k = 100$, capturing potential interactions and social

[3]Our results are robust to correcting for issues related to variation in population density across areas. We first show, descriptively, that we capture similarly sized neighborhoods within similarly large areas, regardless of area density. We further show that the age at arrival results hold when we weight the regressions to account for population density.

networks.

In addition, we focus our analysis on refugees. While it is a group that is heavily understudied, due in large part to data limitations, it is also a group whose integration process may differ from that of other types of immigrants (see, e.g., the discussion in Brell et al., 2020). Moreover, the few existing papers primarily focus on labor market integration (see, for example, Fasani et al., 2018, Battisti et al., 2019 and Dahlberg et al., 2020). By looking at residential integration, our paper is one of the first in this nascent literature to focus on other forms of integration than the labor market. Integration is a multidimensional process, and understanding how it unfolds along these multiple dimensions is important for developing adequate policy responses (see Harder et al., 2018 for the development of a multidimensional integration index and Aksoy et al., 2020 for an application of that index using German data).

We also contribute to the age at arrival literature. Apart from putting refugees center stage, we add to the literature by having a specific focus on residential integration at a small geographical scale. The earlier literature has mostly focused on a range of other outcomes, from education and earnings (Böhlmark 2008, Hermansen 2017, Alexander and Ward 2018, Lemmermann and Riphahn 2018, Ansala et al. 2019), to health (Van den Berg et al. 2014) and social integration (Åslund et al. 2015).

The only earlier paper we know of that has examined the effects of age at arrival on residential integration is Åslund et al. (2015). Our paper differs in two important ways from Åslund et al. (2015). While we focus on recent cohorts of refugees, Åslund et al. (2015) study the children of earlier cohorts of labor immigrants (mainly immigrating from the other Nordic countries or non-Nordic European countries). Ex ante, it is not clear that the effects should be the same for these two vastly different groups of immigrants. In addition, since we use coordinate-based data, we do not have to rely on administratively defined neighborhoods, but can construct individualized neighborhoods.

The paper is organized as follows: in Section 2, we describe the data and elaborate on the k-nearest neighbors approach. In Section 3, we introduce the empirical specification and discuss potential threats to identification. We present and discuss the results from the baseline estimates and the decomposition analysis in Section 4. We conclude in Section 5.

2 Sample selection and neighborhood definition

2.1 Data

The analysis uses Swedish geo-coded register data from the GeoSweden database, which contains information on all residents in Sweden. The data is collected on a yearly basis from 1990 to 2014 and consists of variables from the population and tax registers. Importantly for our study, it contains information on the country of birth, reason for and year of immigration. It additionally includes detailed geographic information on residential location, given by coordinates on a 100×100 meter-level.

Our sample consists of refugee children born between 1974 and 1984 and whose age upon arrival in Sweden is between zero and fifteen.[4] A child is considered a refugee if they either have at least one parent classified as a refugee or their own permit is a refugee permit. We study residential characteristics at age 30, hence we are implicitly restricting to those immigrants who do not return to their home country before that age. For each child, we link information on their own education level, their income (measured in percentile ranks, relative to everyone in their birth cohort), number of siblings, as well as their parents' education and income rank.[5]

[4]The age at arrival variable comes primarily from the in-migration register, which is available from 1990 to 2014. For those arriving before 1990, we use a variable from the income register (Louise) that gives the latest year of immigration. We take the value of this variable when the child first enters the Louise register, at age 16. The earliest cohort that we can observe at age 16 is born in 1974, whereas the youngest cohort we can observe at age 30 is born in 1984. Hence, these data restrictions inform our choice of the cohorts under study.

[5]We measure parents' income rank when the child is between 15 and 19, in order to obtain a measure of financial resources available to the child when they were growing up.

Table 1: Summary statistics

	Mean	Std. dev.	No. of obs.
Siblings sample			
Child percentile income rank	40.16	30.48	22,312
Child has college or above	35.91	n/a	22,137
Parent percentile income rank	12.99	16.30	22,312
At least one parent with college or above	31.78	n/a	21,242
Average age at arrival	9.76	3.41	22,312
Full sample			
Child percentile income rank	41.34	30.77	35,535
Child has college or above	38.69	n/a	35,262
Parent percentile income rank	14.51	17.56	35,535
At least one parent with college or above	35.60	n/a	33,910
Average age at arrival	9.84	3.46	35,535

Notes: Child percentile income rank refers to the position in the earnings distribution relative to everyone in a given cohort. Parents are ranked relative to all parents with children in a given cohort. The earnings measure captures income from employment and self-employment. College or above is defined as having at least a post-secondary education that takes fewer than 3 years to complete.

Table 1 shows summary characteristics for the refugee children in our sample. Since our empirical strategy uses a siblings design, we show how these differ by sample. We see that both the children in the siblings sample and their parents are less likely to have a university degree or above. There are no significant differences in income rank at age 30 in the two groups, and children in both samples arrive, on average, at around age 10.

2.2 Constructing individualized neighborhoods using the k-nearest neighbors approach

The GeoSweden database collects geographical coordinates given on a 100×100 meter level on the 31st of December every year. The 100×100 meter coordinate information in the data allows us to construct individualized neighborhoods of different sizes using the Equipop software developed by Östh (2014).

The procedure for creating individualized neighborhoods is as follows. For each coordinate in the yearly register, we first identify the k-nearest neighbors using the Equipop software, which looks for neighbors in the adjacent 100×100 grids. Similarly, we next identify the neighbors with a particular characteristic among the k-nearest. We then take the ratio of these two values so as to obtain the share of neighbors with a certain characteristic among the k-nearest neighbors for each

coordinate in the yearly register. The k-nearest neighborhood approach ensures that individuals residing at the same coordinate obtain the same value for the share of a certain characteristics among their k-nearest neighbors.

There are various reasons for using the k- nearest neighbor approach. While administrative units differ in size across municipalities, the k-nearest neighbor approach allows us to construct neighborhoods with almost constant counts of individuals, regardless of municipality area and/or size (Östh, 2014; Johnston and Pattie, 2011). Furthermore, with this approach, we can better capture what refugees identify as their neighborhoods, as refugees are placed at the center of their own neighborhoods. Thus, the resulting neighborhood characteristics are good representations of the actual urban context surrounding the individual. Additionally, the k-nearest neighbor approach enables the creation of small neighbourhoods. The small scale analysis, down to $k = 100$, used in this paper captures the characteristics of individuals that refugees are most likely to interact with and potentially be influenced by.

This paper shows results for neighborhoods of two different sizes: 100 and 1000. The different neighborhood sizes allow us to capture the characteristics of individuals that refugees may encounter and possibly interact with both very locally (such as in the building they live in) and more broadly in the area they live in (at work, in shops etc.).

As described above, the algorithm looks for the closest k neighbors, starting from adjacent grids. Variation in density across areas may pose concerns regarding the kinds of neighborhoods we can capture with this procedure. In high-density areas, on the one hand, it can happen that the adjacent grid contains more than k neighbors. In that case, the algorithm reports all the neighbors that are close. In Figure A.1, we show that for most of our sample, the difference between k and the actual number of neighbors that the algorithm finds is between 0 and 200, for both $k = 100$ and $k = 1000$. In low-density areas, by contrast, the algorithm may have to travel to farther grids in order to reach the desired level of k. Figure A.2 shows that this does not seem to be a concern in the case of $k = 100$. As it is expected, slightly larger distances have to be covered in order to reach $k = 1000$ neighbors. Nonetheless, these distances are rarely larger than 400 meters. Together, these figures suggest that we capture similarly sized neighborhoods within similarly large areas, regardless of area density.

We focus on four neighborhood-level characteristics: i) share of natives, where natives are defined as those born in Sweden regardless of their parents' country of birth; ii) share of highly-educated, where high education is defined as having at least

some tertiary education; iii) share of high-earners, that is, those earning above the median in the municipality earnings distribution and iv) share who receive social assistance benefits.[6]

Table 2 shows neighborhood characteristics at age 30 for three different subgroups: natives (column 1), the full sample of refugees (column 2), and the siblings sample of refugees. While the neighborhood characteristics at age 30 of the two refugee groups are very similar to each other, there are some clear differences between the native-born individuals and those arriving as refugees. The two groups differ the least in terms of the share of high-educated neighbors, but the native-born individuals have a larger share of natives, a larger share of high-income earners, and a lower share of individuals on welfare among their neighbors than refugees.

Table 2: Outcomes in different groups

	Natives	Refugees, full sample	Refugees, siblings sample
k = 100			
Share natives	0.86	0.65	0.64
Share high-educated	0.32	0.30	0.29
Share high-earners	0.53	0.46	0.45
Share on welfare	0.04	0.10	0.11
k = 1000			
Share natives	0.85	0.66	0.66
Share high-educated	0.32	0.30	0.30
Share high-earners	0.51	0.45	0.45
Share on welfare	0.04	0.09	0.10
Observations	819,420	35,535	22,312

Notes: Natives are born in Sweden to Swedish parents. Refugees are born abroad to foreign-born parents and arrive in Sweden between the ages of 0 and 15.

3 Empirical strategy

As highlighted by Alexander and Ward (2018), there are two main empirical issues that we have to consider when estimating the effects of age at arrival on neighborhood integration: collinearity and selection bias. In this section, we describe how we address each of these issues in order to get closer to estimating the causal effect of age at arrival on residential integration.

[6]Note that anyone that receives a non-zero amount of social assistance in a given year is considered to be a welfare recipient.

We cannot simultaneously estimate the effect of age at arrival, birth cohort and years spent in Sweden since they are collinear with each other. Therefore, we use natives to identify the birth-cohort neighborhood profile and estimate whether age at arrival influences deviations from this profile. This is accomplished through a two-stage procedure.[7] In the first stage, we use all natives born in the same birth cohorts as the refugees in our sample and estimate the following equation to identify the birth-cohort neighborhood profile of natives:

$$y_i^{native} = \lambda_b + \varepsilon_i \tag{1}$$

where y_i^{native} denotes the natives' neighborhood characteristics and λ_b constitute a full set of birth cohort fixed effects.

In the second step, we use our sample of refugees and examine whether their age at arrival is related to deviations from the native birth-cohort neighborhood profile. This is achieved by regressing immigrants' neighborhood characteristics at age 30 in deviations from the average neighborhood characteristics of natives born in the same birth cohort, estimated in equation (1), $(y_i - \widehat{\lambda_b})$, on age at arrival in Sweden (a_i) and individual and family characteristics that can be observed in the data $(\mathbf{X}_i)$:[8]

$$y_i - \widehat{\lambda_b} = \alpha + \sum_{a=4}^{15} \beta_a I(a_i = a) + \gamma \mathbf{X}_i + \eta_i \tag{2}$$

The second issue we have to address is potential selection bias. The concern is that parents with better unobservables (in terms of, e.g. motivation, parenting skills, and other variables that might be correlated with the outcome variables but that are not observed in the data) may migrate to a larger extent when their children are young. In other words, the controls in equation (2) may not capture all child and parent characteristics that drive both earlier arrival in Sweden and later-life outcomes. We therefore estimate a model with family fixed effects that allows us to identify the effect of every additional year of childhood spent in Sweden on later-life outcomes by using within-family differences in age at arrival. The final model is

[7] This procedure has been used earlier in the literature. See, for example, Alexander and Ward (2018), who apply the procedure in an analysis of the effects of age at arrival during the Age of Mass Migration in the United States on labor market outcomes. We adjust this procedure to our setting and estimate birth-cohort instead of life-cycle profiles since all individuals in our sample are observed at the same age.

[8] The reference category pools ages 0-3.

hence given by:

$$y_{ij} - \widehat{\lambda}_b = \alpha + \sum_{a=4}^{15} \beta_a I(a_{ij} = a) + \mu \text{first-born}_{ij} + \theta \text{female}_{ij} + \phi_j + \eta_{ij} \tag{3}$$

where y_{ij} is the outcome of child i in family j, a_{ij} is the child's age at arrival in Sweden, and ϕ_j is the family fixed effect that captures unobserved family characteristics that are common to all siblings in the same family and constant over time. We follow previous literature that highlights the importance of birth order effects and add a dummy for first-born children (Böhlmark, 2008). We additionally control for gender to capture gender differences in the outcomes we consider.

To get a sense of how the baseline category (i.e., those that arrive at age 0-3) is doing relative to the corresponding cohort of natives, Table 3 reports the mean of the variable $y_i - \widehat{\lambda}_b$ for that age group. It can first be noted that, on average, those that arrive at age 0-3 have approximately a 20 percentage points lower share of natives among their closest neighbors at age 30 compared to their corresponding native cohort. Even though the two groups have been living in Sweden for more or less their whole life, their close neighborhoods are markedly different in ethnic composition.

For the three socio-economic variables, we see a different picture with almost no, or very small differences, between the two groups. At age 30, those that arrived at age 0-3 have 1 percent more high-educated individuals among their closest neighbors, 4 percent more individuals on welfare, and approximately 4 percent fewer high-income earners.

In the bottom panel of Table 3, we also note that, compared to the corresponding native cohort, those that arrive early are 10 percentile ranks lower in terms of earned income, they have half a year of less education, they are 9 percent less likely to be married, and they are 48 percent less likely to be married to a native Swede (conditional on being married at age 30).

4 Results

Our results are presented in the following three sections. In section 4.1, we first present the effects of age at immigration on residential integration, which constitute our baseline estimates. In order to examine the extent to which the effects on residential integration work via other integration channels (income, educational attainment, and intermarriage), we first estimate the effects of age at arrival on

Table 3: Baseline means

	Baseline mean
Panel A: Residential integration outcomes	
Share natives	
$k = 100$	-0.20
$k = 1000$	-0.19
Share high-educated	
$k = 100$	0.01
$k = 1000$	0.01
Share high-earners	
$k = 100$	-0.05
$k = 1000$	-0.04
Share on welfare	
$k = 100$	0.04
$k = 1000$	0.04
Panel B: Other integration outcomes	
Income rank	-10.26
Years of education	-0.57
Marriage	-0.09
Intermarriage	-0.48

Notes: The baseline means refer to the pooled category of those who arrive between the ages of 0 and 3.

these three outcomes in section 4.2 and then decompose the main effect estimated in section 4.1 into the different parts in section 4.3.

4.1 Effects on residential integration

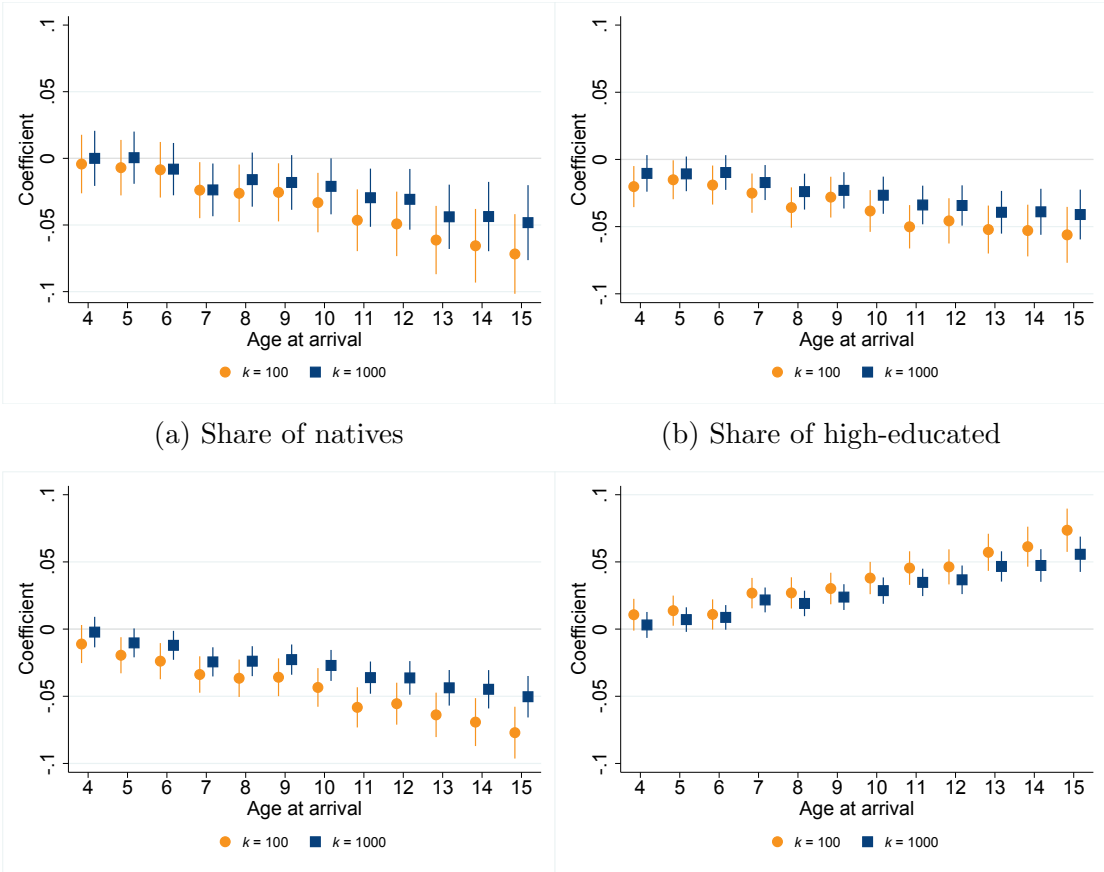
Residential integration in terms of ethnicity

Figure 2a plots the β_a coefficients obtained when estimating equation (3) with share of natives as the dependent variable. We see from the figure that there is a strong negative relationship between age at arrival and the share of natives among the k-nearest neighbors at age 30, for both $k = 100$ and $k = 1000$. The effect is a precisely estimated zero until the age of seven (which roughly corresponds to the school-starting age in Sweden), at which point the effect turns negative.[9] The point estimates at $k = 100$ are slightly more negative than the point estimates at

[9]The effect is relative to those arriving at age 0-3 relative to the corresponding native cohort; c.f. equation (3).

$k = 1000$, implying that refugees have a smaller share of natives among their very closest neighbors.[10] In terms of magnitudes, these coefficients imply that those arriving at age 15 end up in neighborhoods with an approximately five ($k = 1000$) to seven ($k = 100$) percentage point lower share of natives among their neighbors compared to those arriving at ages 0-3 and relative to the corresponding native cohort. These effects amount to 25-35 percent of the mean value for the reference group (c.f. Table 3).

Figure 2: Effect of age at arrival on residential integration outcomes



(a) Share of natives

(b) Share of high-educated

(c) Share of high-earners

(d) Share on welfare

Notes: The figure shows the β_a coefficients obtained when estimating equation (3) and their corresponding 95% confidence intervals.
Source: Own calculations on data from the GeoSweden database.

[10]Our results for the effect of refugees' age at arrival on the share of natives among $k = 1000$ closest neighbors are in line with the results in Åslund et al. (2015) for immigrants in earlier cohorts that typically did not arrive as refugees. They measure residential integration at an administratively-determined unit, the SAMS area, which has on average approximately 1000 inhabitants.

Residential integration in terms of socio-economic characteristics

We focus on three variables when examining the socio-economic composition of the refugee children's neighbors when they reach the age of 30: share highly-educated, share high-income earners, and share on welfare. From the results, presented in Figures 2b-2d, there are three main conclusions that can be drawn.

First, the older a child is when arriving in Sweden, the more disadvantageous their neighborhood at age 30 is (in terms of the neighbors' socio-economic characteristics): there are significantly lower shares of highly educated individuals and high-income earners and a significantly higher share of individuals on welfare compared to the reference category. For example, being older than 10 years old instead of 0-3 years old when immigrating to Sweden implies an approximately 5 percentage points lower share of highly educated neighbors (c.f. Figure 2b), an approximately 6-8 percentage point lower share of high-income earners (c.f. Figure 2c), and an approximately 3-6 percentage point higher share of welfare recipients (c.f. Figure 2d). Comparing these estimates with the mean values for the reference group (see Table 3), we see that the magnitudes of the estimates are sizeable.[11]

Second, as seen for the share of natives, the effect starts being more pronounced at around school-starting age and, in absolute values, the effects seem to continuously increase in magnitude with each age of arrival.

Third, the effects seem to be fairly similar no matter the size of the neighborhood, even though the effects seem to be somewhat less positive for close neighborhoods ($k = 100$).

Since the k-nearest neighbor algorithm might have to search over a long distance to find the nearest neighbors in low density areas, we examine if our results are robust to correcting for variation in population density through weighted regressions. The weight is given by 1/distance, where distance is the number of meters covered to find the desired k-level. As shown in Figure B.1 in the Appendix, the overall pattern is very similar.[12]

[11] In terms of the socio-economic characteristics examined in this paper, those that arrive at age 0-3 live in neighborhoods that are very similar to their native counterparts. This group has a 1 percentage point higher (5 percentage points lower/4 percentage point higher) share of high-educated (high-earners/on welfare) neighbors than the natives.

[12] If the algorithm travels a distance of 0 meters, the weight is simply 1.

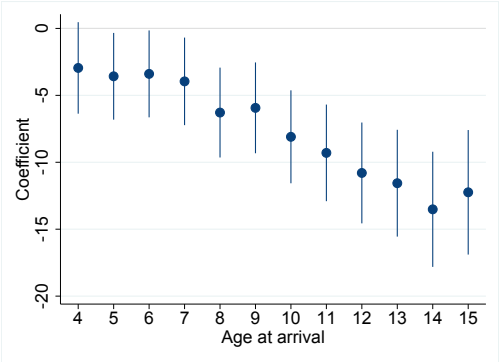
4.2 Effects on labor market, educational, and marital integration

The earlier refugee children arrive in a new country, the more time they have to build country-specific knowledge (e.g. different types of networks, language, cultural habits, institutional knowledge). This country-specific knowledge might also affect other forms of (integration) outcomes that, in turn, might affect residential integration. Here we examine the effects on three other important margins: labor market, educational, and social (marital) integration.

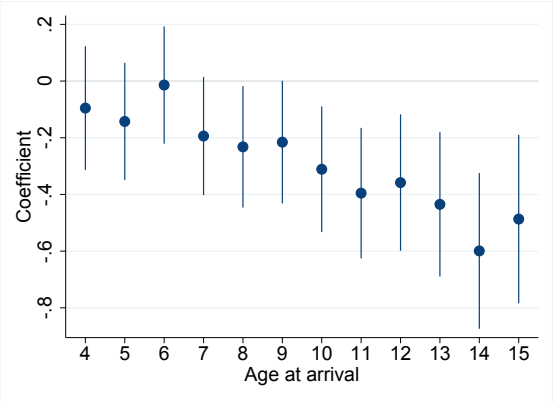
Earlier research on the effects of age at immigration for immigrants in general has found that the earlier they arrive, the better they do on the labor market, the higher their educational achievement is, and the more they marry across ethnic lines. From Figures 3a-3d, we see that this is also true for refugees. For instance, arriving in Sweden at age 15 instead of at age 0-3 implies that refugees have, on average, approximately a 12.5 lower percentile rank in the earnings distribution at age 30, a half a year less of education, and a 28 percentage point lower probability of being married to a native-born partner (conditional on being married at age 30; overall, they are more likely to be married at age 30). Relating the point estimates to the baseline means (see Table 3), the effects are very large.

Given that age at arrival matters for labor market, education, and intermarriage outcomes, it is of interest to examine how much of the baseline estimates of age at arrival on residential integration can be explained by these three intermediate channels. We turn to this in the next section.

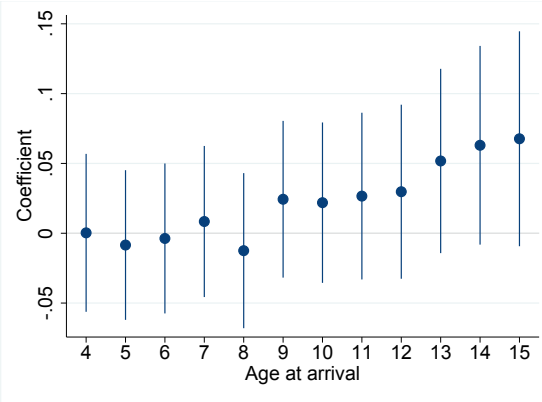
Figure 3: Effect of age at arrival on other integration outcomes



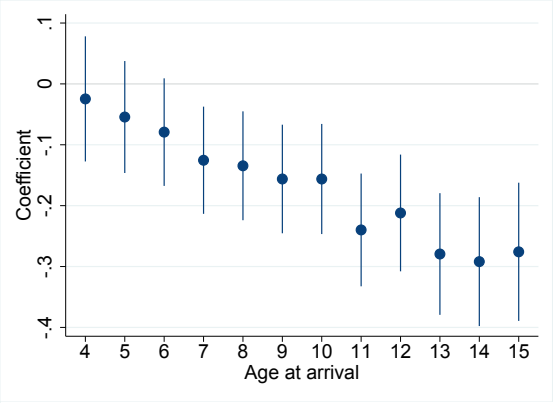
(a) Income rank



(b) Years of education



(c) Married or cohabiting



(d) Married to native, conditional on being married

Notes: The figure shows the β_a coefficients obtained when estimating equation (3) and their corresponding 95% confidence intervals.
Source: Own calculations on data from the GeoSweden database.

4.3 Decomposing the main effect on residential integration

We conduct a decomposition of the main effects in the style of Heckman et al. (2013), where we decompose the effects of age at immigration on neighborhood integration via economic integration, educational integration and intermarriage. To be able to interpret this as a causal effect of the mediators, we need to make strong assumptions. The assumption is that all unobserved factors should be uncorrelated with both age at arrival and the mediators, and orthogonal to the link between the mediators and neighborhood integration. For this reason, we rather think of this method as a descriptive tool to better understand our results.

Since the estimated effects observed in Figures 2a-2d are fairly linear, we have chosen to estimate equation (3) with age of the child entering linearly in the de-

composition exercise (that is, we decompose a linear effect of age at arrival). The reason for this choice is in terms of clarity; instead of presenting a decomposition analysis for each and every age coefficient estimated in Figures 2a-2d, we present an overall decomposition analysis.

The decomposition is conducted in three steps:

1. We first estimate equation (3) with a linear age variable and with the variables income rank, years of education and intermarriage as additional covariates, and save the coefficients on these three additional variables and the main effect of age. These coefficients are in columns (1)-(4) in Table C.1.

2. We then estimate equation (3) with a linear age at arrival variable, separately for each of the variables income rank, years of education and intermarriage as outcome variables. We save the coefficient on the age variable from each of these regressions (columns (5)-(7) in Table C.1).

3. Finally, we calculate the contribution of each of the three "channel" variables. This is done by multiplying the coefficient on each variable as estimated in the first step with the respective coefficient on age as estimated in the second step. This means that we weight the contribution of each variable to the main outcome by the effect of age on that variable. These estimated contributions can be found in columns (8)-(10) of Table C.1.

The total effect is equal to the main effect of age plus the contributions considered, and the shares are equal to each contribution divided by the total effect. These shares are presented in Table 4.[13]

[13]The decomposition presented in Table 4 is based on those individuals that had married at age 30. The reason for this is that we want to decompose the main effects into all three intermediate channels. However, it can be noted that when we use the full sample and decompose the baseline effects into the labor market and education channels, we get shares for these intermediate channels that are very similar to those in Table 4, see Table C.2 and the corresponding Table C.3 with the estimates obtained at steps 1-3 in the decomposition exercise.

Table 4: Decomposition

	Income rank	Years of education	Intermarriage	Residual
Panel A: k = 100				
Share natives	0.0822	0.0234	0.2872	0.6073
Share high-educated	0.0695	0.0890	0.2606	0.5809
Share high-earners	0.0987	0.0276	0.2307	0.6431
Share on welfare	0.0673	0.0222	0.1267	0.7838
Panel B: k = 1000				
Share natives	0.1062	0.0291	0.4015	0.4632
Share high-educated	0.0642	0.0950	0.3065	0.5342
Share high-earners	0.0750	0.0312	0.2305	0.6633
Share on welfare	0.0657	0.0220	0.1588	0.7535

Notes: The table presents the decomposition analysis for the married sample. The estimates to construct this table can be seen in Table C.1 from the Appendix.

The overall impression from the results is that there is a large part of the variation in the baseline effect of age at immigration on neighborhood integration that is still unexplained even after accounting for potential effects going through the three mediators. If we look at $k = 100$, we see that the unexplained variation varies from just below 60 percent (for share high-educated among the $k = 100$ nearest neighbors) to 78 percent (for share on welfare). Of the three mediators, the largest part of the baseline estimates are accounted for by intermarriage and the smallest part by years of education. If we compare the results for $k = 100$ and $k = 1000$, we note that there is a larger unexplained variation in the share of natives for $k = 100$. For the socio-economic variables, the unexplained variation is more similar over neighborhoods of different sizes. It is however worth stressing that the estimated shares presented in Table 4 can probably not be given a causal interpretation, so they should be interpreted with this in mind.

5 Conclusions

The aim of this paper was to examine if, and to what extent, refugee children who arrive at earlier ages in Sweden live in better neighborhoods in adulthood. We reach three conclusions. In our baseline results, we find that those that arrive at younger ages in Sweden (and in particular before school-starting age) are more geographically integrated at age 30: among their very closest neighbors, they have larger shares of natives, highly educated and high-earning individuals and a lower

share of individuals on welfare (compared to their older siblings and once we account for time-invariant, unobserved family characteristics). This indicates that a longer exposure to the host country from an earlier age results in better residential integration outcomes in adulthood, in terms of close neighbors' ethnicity and socio-economic composition.

A long exposure to the host country might, however, also affect other margins, such as labor market and education outcomes, as well as marriage patterns which, in turn, might affect the refugees' choice of residential area at age 30. Examining this, we find that the younger the refugees are when they arrive, the more they earn, the more educated they are and the more likely they are to marry Swedish-born partners by the age of 30.

When examining how large a share of the baseline results is explained by the three intermediate channels, our results indicate that they explain some but far from all of the mean age at arrival effects estimated in the baseline analysis. The unexplained variation, that is, the variation left after accounting for intermediate effects via the labor market, education, and intermarriage channels, is for almost all characteristics and neighborhood sizes larger than 50 percent, and when looking at the characteristics among the 100 nearest neighbors, it varies between 60 and 80 percent.

How can we understand this large unexplained variation in residential integration? What can affect residential integration that does not work via the three intermediate channels examined in this paper? As we see it, there are at least three potential candidates. First, there can be a taste-based explanation that works independently from the three channels. Arriving at different ages can, for example, have differential effects on preferences for certain types of neighbors to interact with. That this might be a possible explanation is indicated by the mean values for those that arrive between ages 0 and 3 (c.f. Table 3); at age 30, those individuals live in neighborhoods that are more or less identical to their corresponding cohort of native-born individuals in terms of socio-economic characteristics, but with markedly fewer individuals born in Sweden. One interpretation for this is that they have preferences for interacting with neighbors that are similar to them, both in terms of socio-economic characteristics and in terms of country of birth.

Second, even if they perform well in the labor market, and can afford to live in any neighborhood that matches their preferences, they may not be able to realize those choices in the presence of discrimination in the housing market. Ethnic-based housing market discrimination could explain the pattern observed in Table 3.

Third, even those that arrive late and do not manage as well in the labor market

may enter more affluent neighborhoods due to the way Swedish housing policies are designed. Tenure mix policies, where the aim is to build different forms of housing tenures in the same neighborhood (e.g. owner-occupied as well as rentals), are intended to promote social mix. In addition, the Swedish rental system is such that rents are not market-determined and individuals are placed in municipality-specific queues for rental apartments, whereby available apartments are offered to the person that has spent the longest in the queue. Since the municipality-owned companies have their properties in all types of areas, affluent as well as less affluent, a given individual can end up in areas with affluent neighbors, independent of their income. We think these types of housing policies have the potential to explain a large part of the unexplained variation observed in the estimates. Examining these three types of explanations would be an important topic for future research.

Acknowledgments

We thank Henrik Andersson, Yvonne Giesing, Stefano Lombardi, Panu Poutvaara and Janne Tukiainen for helpful comments, as well as participants at the 2020 Finnish Economic Association meeting in Tampere and Urban Lab seminar participants at Uppsala University. This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

References

Aksoy, C. G., P. Poutvaara, and F. Schikora (2020): "First Time Around: Local Conditions and Multi-dimensional Integration of Refugees," *IZA Discussion Paper Series*, No. 13914.

Alexander, R. and Z. Ward (2018): "Age at arrival and assimilation during the age of mass migration," *The Journal of Economic History*, 78, 904–937.

Ansala, L., U. Hämäläinen, and M. Sarvimäki (2019): "Age at arrival, parents and neighborhoods: understanding the educational attainment of immigrants' children," *Journal of Economic Geography.*

Åslund, O., A. Böhlmark, and O. N. Skans (2015): "Childhood and family experiences and the social integration of young migrants," *Labour Economics*, 35, 135–144.

Battisti, M., Y. Giesing, and N. Laurentsyeva (2019): "Can job search assistance improve the labour market integration of refugees? Evidence from a field experiment," *Labour Economics*, 61, 101745.

Böhlmark, A. (2008): "Age at immigration and school performance: A siblings analysis using Swedish register data," *Labour Economics*, 15, 1366–1387.

Borjas, G. (1995): "Ethnicity, Neighborhoods, and Human-Capital Externalities," *American Economic Review*, 85, 365–90.

Brell, C., C. Dustmann, and I. Preston (2020): "The Labor Market Integration of Refugee Migrants in High-Income Countries," *Journal of Economic Perspectives*, 34, 94–121.

Chetty, R., N. Hendren, M. R. Jones, and S. R. Porter (2020): "Race and economic opportunity in the United States: An intergenerational perspective," *The Quarterly Journal of Economics*, 135, 711–783.

Chetty, R., N. Hendren, and L. F. Katz (2016): "The effects of exposure to better neighborhoods on children: New evidence from the Moving to Opportunity experiment," *American Economic Review*, 106, 855–902.

Dahlberg, M., J. Egebark, U. Vikman, G. Özcan, et al. (2020): "Labor Market Integration of Low-Educated Refugees: RCT Evidence from an Ambitious Integration Program in Sweden," *IFAU WP*, 21.

De la Roca, J., I. G. Ellen, and J. Steil (2018): "Does segregation matter for Latinos?" *Journal of Housing Economics*, 40, 129–141.

Dinesen, P. T. and K. M. Sønderskov (2018): "Ethnic diversity and social trust: A critical review of the literature and suggestions for a research agenda," *The Oxford Handbook of Social and Political Trust*, 175–204.

Ellen, I. G. and M. A. Turner (1997): "Does neighborhood matter? Assessing recent evidence," *Housing policy debate*, 8, 833–866.

Fasani, F., T. Frattini, and L. Minale (2018): "(The struggle for) Refugee integration into the labour market: Evidence from Europe," *CEPR Discussion Paper No. 12718*.

Graham, B. S. (2018): "Identifying and estimating neighborhood effects," *Journal of Economic Literature*, 56, 450–500.

Harder, N., L. Figueroa, R. M. Gillum, D. Hangartner, D. D. Laitin, and J. Hainmueller (2018): "Multidimensional measure of immigrant integration," *Proceedings of the National Academy of Sciences*, 115, 11483–11488.

Heckman, J., R. Pinto, and P. Savelyev (2013): "Understanding the mechanisms through which an influential early childhood program boosted adult outcomes," *American Economic Review*, 103, 2052–86.

Hermansen, A. S. (2017): "Age at arrival and life chances among childhood immigrants," *Demography*, 54, 201–229.

Johnston, R. and C. Pattie (2011): "Social networks, geography and neighbourhood effects," *The SAGE handbook of social network analysis. SAGE Publications Ltd, London*, 301–311.

Lemmermann, D. and R. T. Riphahn (2018): "The causal effect of age at migration on youth educational attainment," *Economics of Education Review*, 63, 78–99.

List, J. A., F. Momeni, and Y. Zenou (2020): "The social side of early human capital formation: Using a field experiment to estimate the causal impact of neighborhoods," Tech. rep., National Bureau of Economic Research.

Nannestad, P. (2004): "A Game Real Actors Won't Play? Integration of Ethnic Minorities in Denmark as a Collective Action Dilemma," *The International Migration Review*, 38, 287–308.

——— (2009): *Making Integration Work*, Cheltenham, UK: Edward Elgar Publishing.

Östh, J. (2014): "Introducing the EquiPop software," *Department of Social and Economic Geography, Uppsala University: Uppsala, Sweden.*

Sampson, R. J., J. D. Morenoff, and T. Gannon-Rowley (2002): "Assessing "neighborhood effects": Social processes and new directions in research," *Annual review of sociology*, 28, 443–478.

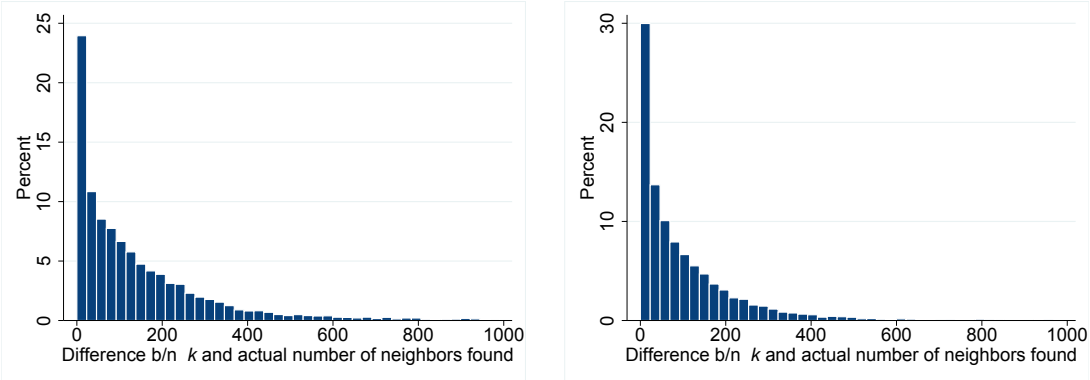
Sharkey, P. and J. W. Faber (2014): "Where, when, why, and for whom do residential contexts matter? Moving away from the dichotomous understanding of neighborhood effects," *Annual review of sociology*, 40, 559–579.

Van den Berg, G. J., P. Lundborg, P. Nystedt, and D.-O. Rooth (2014): "Critical periods during childhood and adolescence," *Journal of the European Economic Association*, 12, 1521–1557.

Appendices

A Checks for k-nearest algorithm

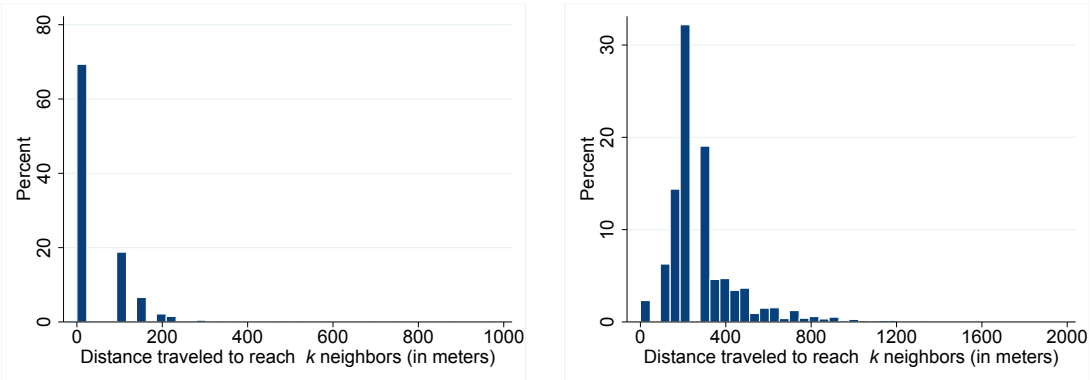
Figure A.1: Difference between k and observed k



(a) $k = 100$ (b) $k = 1000$

Notes: The x-axis is cropped in order to make the graphs easier to read.

Figure A.2: Distance traveled by EquiPop software to find k neighbors

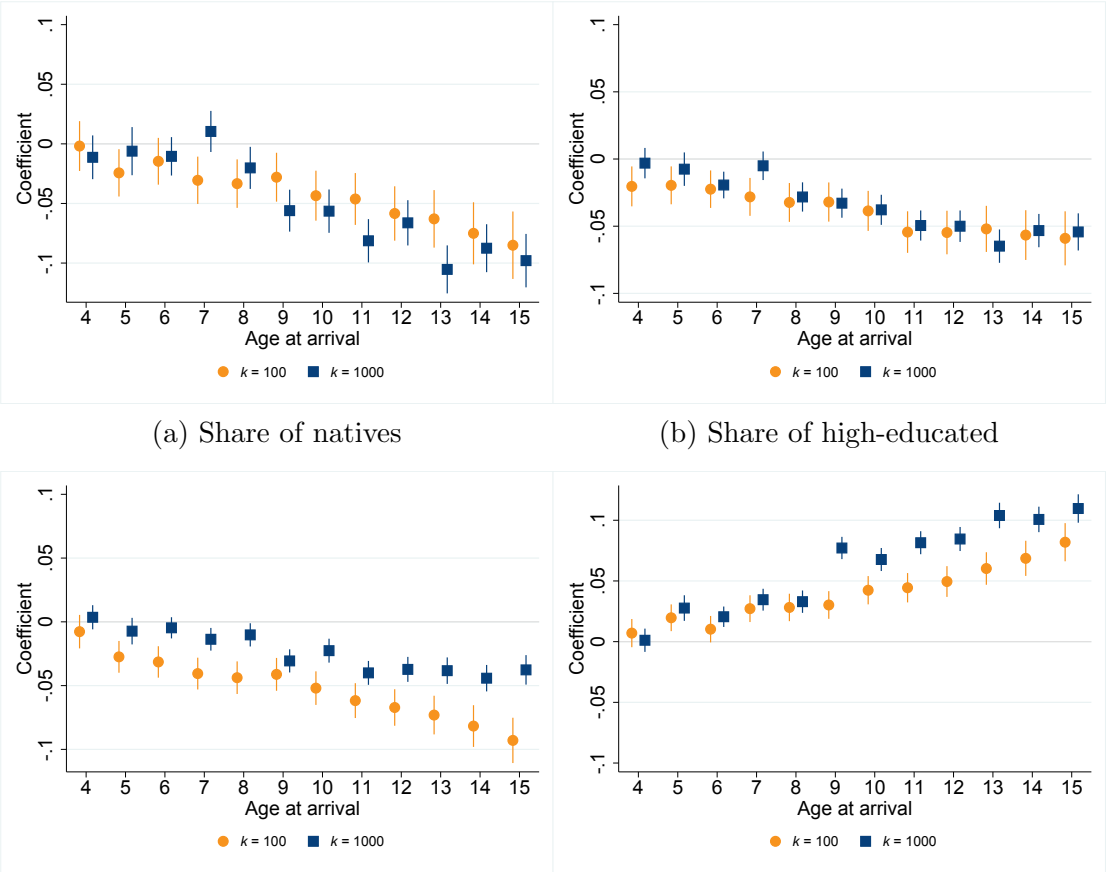


(a) $k = 100$ (b) $k = 1000$

Notes: The x-axis is cropped in order to make the graphs easier to read.

B Results from weighted regressions

Figure B.1: Effect of age at arrival on residential integration outcomes



(a) Share of natives

(b) Share of high-educated

(c) Share of high-earners

(d) Share on welfare

Notes: The figure shows the coefficients and the corresponding 95% confidence intervals from the weighted regressions.
Source: Own calculations on data from the GeoSweden database.

C Decomposition

C.1 Married sample

Table C.1: Decomposition: steps to obtain shares; married sample

	Coefficients from augmented eq. (1)				Effect of age on channels			Contributions				Shares			
	(1) Age	(2) I	(3) ED	(4) IM	(5) I	(6) ED	(7) IM	(8) I (2) × (5)	(9) ED (3) × (6)	(10) IM (4) × (7)	(11) T (1) + (8) +(9) + (10)	(12) I (8)/(11)	(13) ED (9)/(11)	(14) IM (10)/(11)	(15) R (1)/(11)
Panel A: k = 100															
Share natives	-0.0051	0.0007	0.0049	0.0945	-0.9235	-0.0401	-0.0254	-0.0007	-0.0002	-0.0024	-0.0084	0.0822	0.0234	0.2872	0.6073
Share high-educated	-0.0021	0.0003	0.0082	0.0378	-0.9235	-0.0401	-0.0254	-0.0003	-0.0003	-0.0010	-0.0037	0.0695	0.0890	0.2606	0.5809
Share high-earners	-0.0032	0.0005	0.0034	0.0453	-0.9235	-0.0401	-0.0254	-0.0005	-0.0001	-0.0012	-0.0050	0.0987	0.0276	0.2307	0.6431
Share on welfare	0.0053	-0.0005	-0.0037	-0.0335	-0.9235	-0.0401	-0.0254	0.0005	0.0001	0.0009	0.0067	0.0673	0.0222	0.1267	0.7838
Panel A: k = 1000															
Share natives	-0.0023	0.0006	0.0037	0.0799	-0.9235	-0.0401	-0.0254	-0.0005	-0.0001	-0.0020	-0.0051	0.1062	0.0291	0.4015	0.4632
Share high-educated	-0.0015	0.0002	0.0065	0.0332	-0.9235	-0.0401	-0.0254	-0.0002	-0.0003	-0.0008	-0.0028	0.0642	0.0950	0.3065	0.5342
Share high-earners	-0.0028	0.0003	0.0033	0.0389	-0.9235	-0.0401	-0.0254	-0.0003	-0.0001	-0.0010	-0.0043	0.0750	0.0312	0.2305	0.6633
Share on welfare	0.0037	-0.0004	-0.0027	-0.0308	-0.9235	-0.0401	-0.0254	0.0003	0.0001	0.0008	0.0049	0.0657	0.0220	0.1588	0.7535

Notes: The table shows the decomposition analysis for the married sample. The variable I denotes the income rank. ED stands for education level. IM represents intermarriage. Columns (1-3) show the regressions using equation (1). Columns (4-5) shows the estimated effect of age on income rank and years of education. T denotes the total effect and R represents the residual.

C.2 Full sample

Table C.2: Decomposition

	Income rank	Years of education	Residual
Panel A: k = 100			
Share natives	0.1003	0.0350	0.8647
Share high-educated	0.0947	0.1077	0.7977
Share high-earners	0.1128	0.0307	0.8565
Share on welfare	0.0790	0.0295	0.8915
Panel B: k = 1000			
Share natives	0.1165	0.0439	0.8396
Share high-educated	0.0887	0.1080	0.8033
Share high-earners	0.1035	0.0326	0.8640
Share on welfare	0.0684	0.0274	0.9042

Notes: The table presents the decomposition analysis for the full sample. The estimates to construct this table can be seen in Table C.3.

Table C.3: Decomposition: steps to obtain shares; full sample

	Coefficients from augmented eq. (1)			Effect of age on channels		Contributions			Shares		
	(1) Age	(2) I	(3) ED	(4) I	(5) ED	(6) I (2) × (4)	(7) ED (3) × (5)	(8) T (1) + (6) + (7)	(9) I (6)/(8)	(10) ED (7)/(8)	(11) R (1)/(8)
Panel A: k = 100											
Share natives	-0.0055	0.0006	0.0048	-1.0858	-0.0461	-0.0006	-0.0002	-0.0063	0.1003	0.0350	0.8647
Share high-educated	-0.0034	0.0004	0.0100	-1.0858	-0.0461	-0.0004	-0.0005	-0.0043	0.0947	0.1077	0.7977
Share high-earners	-0.0048	0.0006	0.0038	-1.0858	-0.0461	-0.0006	-0.0002	-0.0057	0.1128	0.0307	0.8565
Share on welfare	0.0050	-0.0004	-0.0036	-1.0858	-0.0461	0.0004	0.0002	0.0056	0.0790	0.0295	0.0295
Panel A: k = 1000											
Share natives	-0.0035	0.0004	0.0039	-1.0858	-0.0461	-0.0005	-0.0002	-0.0041	0.1165	0.0439	0.8396
Share high-educated	-0.0027	0.0003	0.0080	-1.0858	-0.0461	-0.0003	-0.0004	-0.0034	0.0887	0.1080	0.8033
Share high-earners	-0.0034	0.0004	0.0027	-1.0858	-0.0461	-0.0004	-0.0001	-0.0039	0.1035	0.0326	0.8640
Share on welfare	0.0041	-0.0003	-0.0027	-1.0858	-0.0461	0.0003	0.0001	0.0046	0.0684	0.0274	0.9042

Notes: The table shows the decomposition analysis for the full sample. The variable I denotes the income rank. ED stands for years of education. Columns (1-3) show the regressions using equation (1). Columns (4-5) shows the estimated effect of age on income rank and years of education. T denotes the total effect and R represents the residual.