# The social context of nearest neighbors shapes educational attainment regardless of class origin

Hedefalk, Finn; Dribe, Martin

*Published in:*
Proceedings of the National Academy of Sciences of the United States of America

*DOI:*
10.1073/pnas.1922532117

2020

*Document Version:*
Publisher's PDF, also known as Version of record

# The social context of nearest neighbors shapes educational attainment regardless of class origin


**Finn Hedefalk[a,b,1] and Martin Dribe[a,b]**

[a]Centre for Economic Demography, Lund University, SE-223 63 Lund, Sweden; and [b]Department of Economic History, Lund University, SE-220 07 Lund, Sweden



Edited by Douglas S. Massey, Princeton University, Princeton, NJ, and approved May 13, 2020 (received for review December 21, 2019)



**We study the association between sociospatial neighborhood conditions throughout childhood and educational attainment in adulthood. Using unique longitudinal microdata for a medium-sized Swedish town, we geocode its population at the address level, 1939 to 1967, and link individuals to national registers, 1968 to 2015. Thus, we adopt a long-term perspective on the importance of nearby neighbors during a period when higher education expanded. Applying a method for estimating individual neighborhoods at the address level, we analyze the association between the geographically weighted social class of the nearest 6 to 100 childhood neighbors (ages 2 to 17), and the likelihood of obtaining a university degree by age 40, controlling for both family social class and school districts. We show that even when growing up in a town with relatively low economic inequality, the social class of the nearest same-age neighbors in childhood was associated with educational attainment, and that the associations were similar regardless of class origin. Growing up in low-class neighborhoods lowered educational attainment; growing up in high-class neighborhoods increased attainment. Social class and neighborhoods reinforced each other, implying that high-class children clustered with each other had much higher odds of obtaining a university degree than low-class children from low-class neighborhoods. Thus, even if all groups benefited from the great expansion of free higher education in Sweden (1960s to 1970s), the large inequalities between the classes and neighborhoods remained unchanged throughout the period. These findings show the importance of an advantageous background, both regarding the immediate family and the networks of nearby people of the same age.**



neighborhood | inequality | educational attainment | microlevel | geography


There is a large amount of literature in economics and sociology arguing that children's opportunities later in life are shaped by the social and physical characteristics of the neighborhoods they grow up in (1–15). In the context of American inner-city poverty, Wilson (16) maintains that neighborhood socioeconomic status (SES) affects aspirations, attitudes, and motivation, which in turn shape, e.g., individual educational outcomes. The mechanism behind this association is the social isolation between different SES groups, and most notably the isolation of low-SES groups in poverty-stricken areas of the large urban centers. Moreover, in a series of papers, Chetty et al. (17–20) use large longitudinal datasets to show the importance of neighborhood economic conditions and upward mobility for children's life chances. Related to these studies, Manduca and Sampson (21) and Donnelly et al. (13) show how harsh social and physical environments directly limit intergenerational social mobility for children, whereas areas with high upward mobility positively influence their cognitive abilities. Other factors within neighborhoods that influence children include concentrated poverty, social control and cohesion, exposure to violence, social and ethnic segregation, and physical properties, such as lead exposure, air pollution, population density, the presence of green spaces, and the overall built environment (22–29). Less visible mechanisms are the social capital within neighborhoods and the possible use of neighbors as important weak ties for one's upward mobility (30–32).

There are different theoretical perspectives aiming to explain the impact of neighborhoods on individual social outcomes including academic achievements, focusing on the influence of peers, adults in the neighborhood, and external adults (e.g., teachers, police, etc.) (15). Peers could influence children's chances of attending, as well as finishing, school through a direct influence on attitudes, norms, and behavior. Especially bad behavior is expected to be contagious in neighborhoods, but good behavior could have similar effects. Nonparental adults in the neighborhood could also affect children's outcomes by acting as role models, especially in showing positive results of good behavior, hard work, and study. In addition, these adults could help maintaining good order in the neighborhoods, thereby controlling adverse behavior. Finally, external adults, such as teachers or law enforcement, could have a role in creating neighborhood effects if low-SES neighborhoods systematically attract lower quality external adults to their institutions. These theories predict that children from low-SES origins benefit from growing up in a high-SES neighborhood. On the contrary, according to resource competition and relative deprivation models, low-SES children in high-SES neighborhoods, or in schools with more high-SES students, may suffer from greater grade competition and higher perception of own deprivation. This could lower their chances of

SOCIAL SCIENCES

## Significance

**Much neighborhood research has focused on contemporary and segregated cities in the United States, but less on small and more homogenous cities. Additionally, neighborhood conditions are often estimated using administrative borders, which bias results. We adopt a long-term perspective on the importance of childhood neighbors, using more realistic methods of neighborhood conditions. We estimate individual neighborhoods at the address level, using geocoded longitudinal microdata (1939 to 2015) for a medium-sized Swedish town. We show that even when growing up in an economically relatively equal population, when higher education expanded greatly, the social class of the nearest childhood neighbors was important for educational achievements, regardless of social class and schools. Associations are strongest for boys, but with similar patterns across genders.**


Author contributions: F.H. and M.D. designed research; F.H. performed research; F.H. and M.D. analyzed data; and F.H. and M.D. wrote the paper.

The authors declare no competing interest.

This article is a PNAS Direct Submission.





Data deposition: The code and scripts used in the analyses have been deposited in the Harvard Dataverse V2 repository (https://doi.org/10.7910/DVN/XZ9FMI).



[1]To whom correspondence may be addressed. Email: finn.hedefalk@ekh.lu.se.



This article contains supporting information online at https://www.pnas.org/lookup/suppl/doi:10.1073/pnas.1922532117/-/DCSupplemental.

higher educational achievement, or at least counteract other positive effects of growing up in a high-SES neighborhood (33).

In the context of educational attainment, we hypothesize that growing up in a high-SES neighborhood increases the likelihood of obtaining a university degree. Conversely, growing up in a low-SES neighborhood lowers the likelihood of reaching this level of education. Most focus in the literature has been on children from low-SES origins, but neighborhood SES may also affect the educational achievements of children from high-SES origins. Due to the disadvantaged position of children from low-SES origins, neighborhood effects may be stronger for low-SES origins than for high-SES origins. It is difficult to empirically distinguish between these mechanisms, as well as to identify causal effects of neighborhood characteristics on individual outcomes (34, 35).

Despite a considerable literature on how neighborhood affects individual outcomes of children, several areas remain underexplored. Much of the research has focused on the United States, which is a special context given the interaction between race and SES. In addition, there has been an almost exclusive focus in the literature on more recent times, often on heavily segregated and large urban areas. Much of the world's population lives in smaller, more homogenous cites, and the question is how much the neighborhood matters within such areas? In addition, segregation and neighborhood differences within these areas can be detected, but mostly at smaller scales, with the use of detailed geographic data (36, 37). For such cases, we need to use measures of the few closest neighbors to study the importance of neighborhoods.

Although important exceptions exist (14, 38, 39), most longitudinal studies have primarily been limited to estimating neighborhood conditions according to various forms of administrative units (wards, enumeration districts, zip codes). Studies using more detailed spatial estimates of neighborhoods have often done so without a longitudinal component (34, 35). Studies using data aggregated in larger geographic units are plagued by the well-known modifiable areal unit problem and the uncertain geographic context problem (40–42). That is, when using arbitrary and large areal units for deriving neighborhood variables, one misses crucial information on the specific physical and social factors that influence individual behavior. Therefore, serious bias in the predictions of the neighborhood effects may be introduced (42).

We overcome these problems by using longitudinal microdata on a medium-sized Swedish industrial port town (Landskrona), in which we geocode the residential histories of the full population at address level for the entire period 1939 to 1967. Individuals are linked to Swedish national register data for the period 1968 to 2015 using unique personal identifiers, which allows us to study educational attainment in adulthood of children growing up in the city, regardless of where they reside in Sweden. Therefore, we avoid bias from only looking at the stayer population. We study the association between socio-spatial neighborhood conditions throughout childhood and educational choice in adulthood, net of any intergenerational transmission of disadvantage within the family. Economically, the population of Landskrona had relatively equal income; the period-average income-Gini index was 0.27 (*SI Appendix*, Fig. S1). This period also encapsulates the greatest expansion of higher education in Sweden. Thus, within a relatively equal city with free education, we can adopt a long-term perspective on the importance of neighbors during a period of great increase in college attendance.

Moreover, using geocoded data on the address level, we produce fine-scale and geographically weighted measures of the closest neighbors of the same age throughout childhood. In addition, we account for both the social and spatial characteristics of the individual neighborhood, including the elementary school the children most likely attended. Thus, we build upon the work done by others (e.g., refs. 14, 39) and apply a more realistic approach of estimating long-term social influences from the nearest neighbors than previously used measures that have been dependent on administrative borders.

## Materials and Methods

**Data Sample.** The longitudinal and individual-level data (1939 to 1967) for the city of Landskrona come from the Scanian Economic–Demographic Database (SEDD) (43). During the post-World War II period, Landskrona was a medium-sized Swedish city with a strong manufacturing and shipyard industry. It had a population of 25,000 in 1949 and ∼27,000 to 30,000 for the period 1960 to 2010. The primary sources for the database are continuous population registers and income and taxation registers. Information on birth, marriages, deaths, and in- and out-migrations has also been linked to the data. The SEDD database has been linked to national longitudinal register data for the period 1968 to 2015 from Statistics Sweden (44, 45). The dataset for analysis was created using a programme developed by Quaranta (46).

We geocoded 98% of the person-time for the ∼77,000 individuals in the historical dataset at the address level, providing the full residential histories of the individuals living in the city (*SI Appendix*, section 1.1). In addition, these address points were linked to the buildings in the city. We also created an object lifeline representation of ∼90% of the buildings and streets in Landskrona. Hence, we have information on when a road and building started and ceased to exist (43). This allows us to use geographical data that are correct for each time point that we study, which is important when estimating individually based neighborhood conditions.

**Research Strategy.** First, we quantify individual, social, and environmental neighborhood variables for children aged 2, 5, 8, 11, 14, and 17 in Landskrona between 1939 and 1967. By doing so, we cover the entire childhood and are able to account for variations in the neighborhood influence at different ages. For example, at age 14, peer influence may affect children's choice of high school orientation (47), whereas other mechanisms may be in play at younger ages. The education of these individuals is later followed up at age 40 in the national registers (1968 to 2015). The total sample of all age groups was 14,436 (*SI Appendix*, section 1.2). The same individual child is on average observed 3.3 times across the age groups.

We focus on two main childhood variables: social class (class origin) and geographically weighted social class of the neighborhood (neighborhood class). At the individual level, we have detailed socioeconomic and demographic information. We measure social class of children based on the occupation of the father and use social class rather than education of the parents to predict educational attainment in adulthood because of the low proportion of adults (i.e., parents) having higher education during the period 1939 to 1967. The great expansion of higher education in Sweden did not take place until the 1960s and early 1970s, and affected cohorts born after 1930 (see, e.g., ref. 48, p. 154; *SI Appendix*, Fig. S2).

Second, the children growing up in Landskrona from 1939 to 1967 are followed until age 40, regardless of where they live within Sweden. For these individuals, we estimate logistic regression models to analyze the association between childhood neighborhood conditions and the likelihood of obtaining a 3-y university degree by age 40. The variables used in the analyses are categorized into four model groups: individual variables, neighborhood class, social environment, and physical environment. These are presented as follows.

**Dependent Variable.**

***Education level.*** The binary outcome variable, which is based on information on the highest level of education reached at age 40 (an age when most people have finished their education in Sweden). This information is coded according to the national standard Swedish education nomenclature (SUN) for classification of education. A value of 1 indicates that the individual has completed a 3-y university degree or higher.

**Class Origin and Other Individual Variables.**

***Class origin.*** We measure socioeconomic status based on the father's occupation. Data on occupation are obtained from several sources: demographic events, population registers, and annual data from the income registers. Occupational notations are coded in an internationally comparable coding scheme for historical occupations (HISCO) (49) and then grouped into HISCLASS, a 12-category social class scheme based on skill level, degree of supervision, whether manual or nonmanual, and whether urban or rural (50). We define three classes: high class (HISCLASS: 1 to 5), midclass (HISCLASS 6 to 7), and low class (HISCLASS 8 to 12) (see *SI Appendix*, section 1.3, for more information). The high-class group have mostly white-collar occupations such as managers, professionals, and clerical and sales personnel, whereas the midclass (foremen and medium-skilled workers) and low-class (lower and unskilled workers) groups are mostly blue-collar workers. Other individual variables included as controls are household size, presence of mother, and birth year (see *SI Appendix*, section 1.4, for more information).

**Neighborhood Class Variable.** Using geocoded data at the address level, we estimate social neighborhood conditions based on the *k*-nearest neighbors. The first step is to construct annual matrices containing the shortest Euclidean distances between the individuals. The matrices contain the distances between each child within their age group and their *k*-nearest neighbors of the same age, ±1 y. This interval is used to expand the sample for each neighborhood, capture more children of similar ages, and include information throughout childhood (1 to 18 y). The final dataset contains yearly snapshot information on each individual's neighbors at the end of the year (19xx-12-31) for the period 1939 to 1967. From these matrices, we create individual neighborhoods from the *k*-nearest neighbors that do not live in the household of the individual. We use a range of closest neighbors ($k = 6$, 13, 25, 50, and 100) to 1) capture segregation that may exist in smaller and more homogenous cities (36); 2) include possible peer effects as well as other neighborhood effects; and 3) study how the influences from neighbors change when increasing *k*. Based on the individual neighborhoods, we construct the geographically weighted (GW) *neighborhood class* variable. The variable is based on the class origin share of the *k*-nearest individuals for each specific year. For the *k*-nearest neighbors of individual *i*, we define the geographically weighted share of each class origin *c* as follows:

$$GW\ neighborhood\ class_c = \sum_{j=1}^{j=n}\left(\frac{W_{ij_c}}{\sum_{j=1}^{j=n} W_{ij}}\right), \qquad [1]$$

$$W_{ij} = e^{-0.5\cdot\left(\frac{d_{ij}}{b}\right)^2},$$

where $W_{ij}$ is the spatial weight implemented as a Gaussian distance function between individual *i* and any neighbor *j* (cf. ref. 51), and $j_c$ is the neighbor of the specific class *c*. In the Gaussian distance function, the bandwidth *b* limits the search of the neighbors, and $d_{ij}$ is the Euclidean distance between the address points of individual *i* and neighbor *j*. In this study, we use an adaptive bandwidth, which represents the maximum distance between individual *i* and its *k*-nearest neighbors. Moreover, to account for the variation in population density and uneven distribution of address points in the data, we use the relative spatial weight between individual *i* and neighbor *j* (relative to the spatial weights of all other neighbors). The Gaussian distance function is chosen as it is commonly used when modeling spatial relationships (52). Sensitivity tests are also performed using an exponential distance function as well as equal weights (no distance decay) (*SI Appendix*, section 1.5). Fig. 1 shows an example of the 13 nearest neighbors *j* and their class *c*, of the same age ±1 y to individual *i*. The share of the neighbors' classes, which are five high class, four midclass, and four low class, are weighted according to Eq. **1**. The closest three neighbors, all of different classes, reside only a few meters from the individual. Finally, we categorize the neighborhood class as follows: high class, a *k*-neighborhood with at least a GW share of 50% high-class neighbors, and less than 25% low-class neighbors; low class, a *k*-neighborhood with at least a GW share of 50% low-class neighbors, and less than 25% high-class neighbors; and midclass, a *k*-neighborhood that is not classified as high class or low class.

**Social and Physical Environment Variables.** We control for other social and physical environment aspects within the neighborhood, which may affect the association between neighborhood class and educational attainment (see *SI Appendix*, section 1.4, for more details on these variables, and *SI Appendix*, Fig. S8, for sensitivity tests on school variables). The social environment variables are as follows: Distance to Landskrona's only secondary and high school, elementary school districts, GW sex composition, GW share of neighbors with missing mother, residential mobility, and locally born For the physical environment within the neighborhood, we try to account for aspects such as air pollution, which may affect the cognitive development of children (26, 27). The variables that use information from roads and buildings are time-varying because they account for the construction of new roads and buildings throughout the study period. The physical environment variables are as follows: Proximity to major road, Road density, Building type (apartment block or single house/town house), and Population density.

**Statistical Analysis.** By estimating six logistic regression models, we analyze how neighborhood conditions are associated with education in adulthood (*SI Appendix*, section 1.6). The binary outcome variable of interest is whether the highest level of education reached at age 40 is at least a 3-y university degree. We run separate models for sex, the six age groups (2, 5, 8, 11, 14, and 17) and the number of neighbors (6, 13, 25, 50, and 100). Note that the same child may appear in multiple age groups. We first estimate three basic models, which control only for the individual variables, in addition to class origin (model 1), neighborhood class (model 2), and both class origin and neighborhood class (model 3). Then, we extend model 3 by first adding the social environmental variables (model 4) and thereafter the physical environmental variables (model 5). The last model (model 6) includes an interaction between the variables class origin and neighborhood class, using the settings from model 5. The Bayesian information criterion is used to compare the fit of each estimated model across the neighborhood sizes and spatial weight methods (*SI Appendix*, section 1.5). We also test for spatial autocorrelation of the residuals, since this may introduce bias in the models (*SI Appendix*, section 1.7).

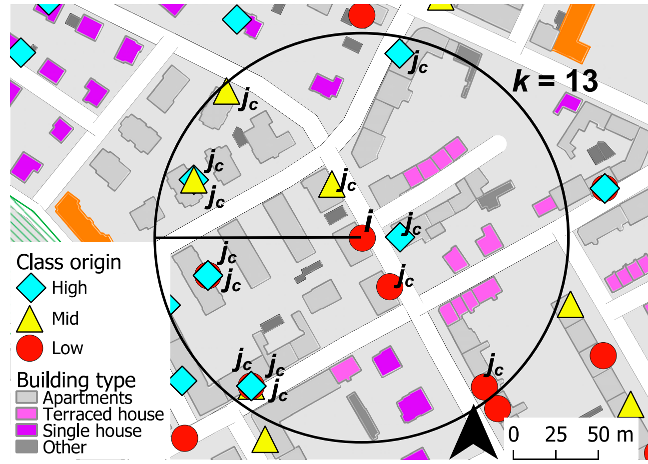


**Fig. 1.** Example of an individual neighborhood for individual *i*, aged 8, containing the 13 nearest neighbors of ages 7 to 9 in Landskrona, 1959-12-31. The background map shows buildings, streets (white lines), and class origin. Note: Some neighbors reside at the same address point.

**Data Availability.** The individual-level data from the SEDD and the Statistics Sweden (43–45) are protected by Swedish personal integrity laws, as well as other regulations and confidentiality restrictions. The analyses are performed on Statistics Sweden's restricted platform MONA (Microdata Online Access), and it is not allowed to openly share these data. However, if relevant permissions have been obtained in accordance with the restrictions stated by the Regional Ethical Review Board, the Swedish Data Inspection Board and Lund University, researchers can gain access to the microdata used in this study. See https://www.ed.lu.se/databases/sedd and https://www.scb.se/en/services/guidance-for-researchers-and-universities for more information on how to access these data. Relevant aggregated data can, however, be shared. In addition, the code and scripts used in the analyses are available at https://dataverse.harvard.edu/dataverse/LPSD/.

## Results

**Descriptive Results.** The great expansion of higher education in Sweden took place in the 1960s and early 1970s, implying that many of the children in Landskrona reached a higher educational level than their parents. All social classes seem to have benefited from this expansion (*SI Appendix*, Fig. S2), but there was a consistently large difference between the high class (primarily white-collar workers) and the mid- and low classes (primarily blue-collar workers) throughout the period. The period average for each class that had obtained at least a 3-y university degree as of age 40 was 25.3% (high), 10.3% (mid), and 7.7% (low).

Fig. 2 shows the distribution of the neighborhood classes, as well as other variables used in the statistical analyses, averaged on a 75 × 75-m grid for the period 1960 to 1967. For descriptive statistics of these and all other variables, see *SI Appendix*, Tables S1 and S2. The maps reveal clusters of the neighborhoods; the high-class neighborhoods are located in the city center near

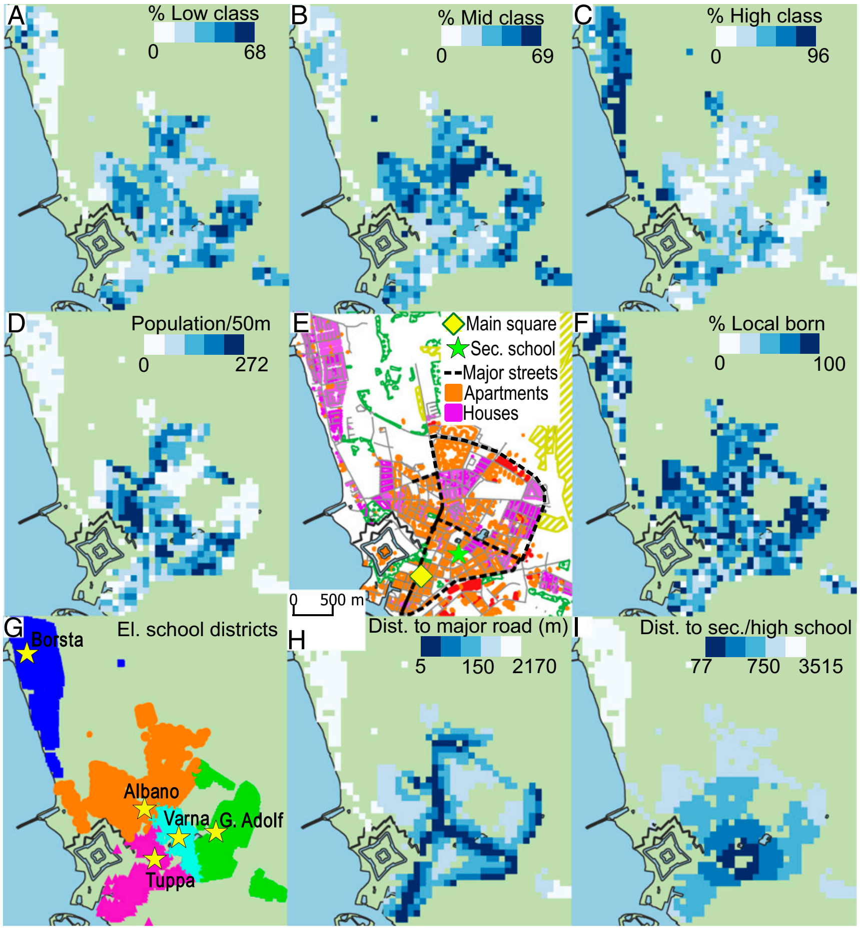


**Fig. 2.** Spatial distribution of neighborhood classes and other variables in Landskrona for all age groups (2–17). The variable values on the individual level for each 13-neighborhood are averaged on a 75 × 75-m grid for the period 1960 to 1967. (*A*) GW % of low-class neighbors; (*B*) GW % of midclass neighbors; (*C*) GW % of high-class neighbors; (*D*) population density (population/50-m buffer from each individual); (*E*) map of Landskrona in 1967 (Sec. school, secondary and high school); (*F*) percentage of locally born children; (*G*) estimated elementary school districts; (*H*) distance to major road (in 1967); and (*I*) distance to secondary and high school.

Landskrona's secondary and high school (residing in apartments) and in the periphery (residing in single house areas), whereas the low-class neighborhoods are located between the center and periphery.

There are also considerable changes in the neighborhood distribution throughout the study period. As shown in Fig. 3, the Getis-Ord Gi* statistics (cf. ref. 53) show several significant clusters of high classes in the city center in 1947, whereas these

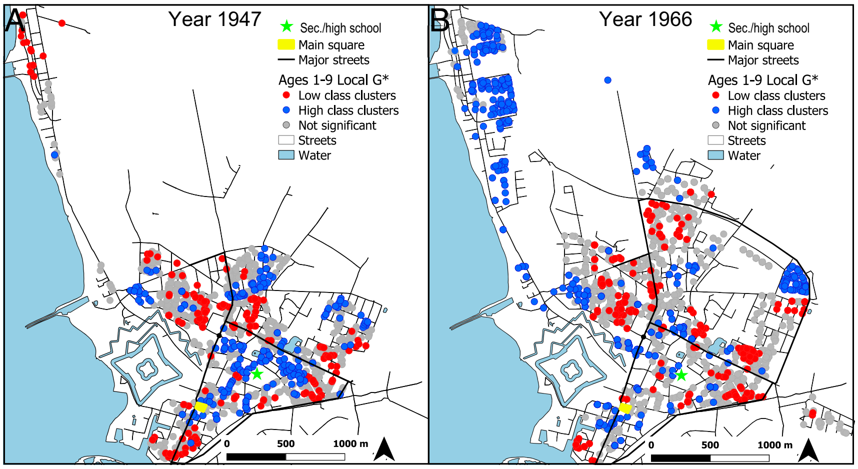


**Fig. 3.** Snapshots of local clusters (Getis-Ord Gi*) in Landskrona (*A*) year 1947, (*B*) year 1966.

clusters have moved to the periphery of the city in 1966. These spatial changes were likely related to the construction of many new buildings in Landskrona from 1950 onward.

**Empirical Results.** Fig. 4 shows the results of five logistic regression models for each age group and sex, using the 13-nearest neighbors' specification (see *SI Appendix*, Figs. S3 and S4, for the results on the full range of $k$). The models estimate the association between childhood neighborhood class and having a university degree at age 40. Only the variables high-class origin and high-class neighborhood (reference classes: low-class origin and low-class neighborhood) are shown in the figure (*SI Appendix*, Fig. S5, includes the full regression outputs). For children residing in midclass neighborhoods, only boys aged 8 were associated with significantly higher odds of obtaining a university degree than children residing in low-class neighborhoods.

The first three models adjust only for birth year, household size, and presence of the mother, in addition to class origin (M1), neighborhood class (M2), and class origin plus neighborhood class (M3). The fourth model (M4) adds social-environmental variables (locally born, residential stability, relative distance to secondary and high school, estimated elementary school district, GW share of female neighbors, and GW share of missing mothers), and the fifth model (M5) adds physical-environmental variables (population density, road density, distance to main road, and type of building).

Throughout the period and for both sexes, there is a strong positive association between high-class origin and the likelihood of having a university degree (Fig. 4 *A* and *B*). The magnitude of this relationship is little affected when controlling for neighborhood class (M3) as well as for the social and physical-environmental variables (M4 and M5). The models also show significant associations with neighborhood class throughout the period (Fig. 4 *C* and *D*), although the association is weaker than that for class origin. The magnitude of the association for neighborhood class decreases most when the basic neighborhood class model (M2) is extended with the inclusion of the class origin variable (M3). However, the association is significant for boys for all models except at age 14; for girls, it is significant only for age 14 (at $P < 0.1$ for ages 8 and 11).

Moreover, the pattern throughout the age groups differs slightly between the sexes: the association between both high-class origin and high-class neighborhood and the odds of having a university degree peaks in the later ages for girls but not for boys. For girls, the strongest influence of both class origin and neighborhood class is observed at age 14. In contrast, boys are not significantly influenced by their nearest neighbors at age 14, but the strongest links are found at ages 5, 8, and 11.

From the comparison of the different sizes of neighborhoods (*SI Appendix*, Figs. S3 and S4), we observe consistent associations of neighborhood class across the sizes of $k$. The strongest associations and best model fits are found at the lower values ($k = 6$, 13, and 25), whereas the associations at $k = 100$ are seldom significant and the models perform worse. The change of the influence from the neighborhood class when altering the number of neighbors differs slightly across age groups and sex. Moreover, the comparisons of the models using Gaussian and exponential weights, as well as when using no weight, indicate that the models using Gaussian distance decay performs slightly better overall than its counterparts, but with a variation across sex and age (*SI Appendix*, Figs. S6 and S7). Last, Moran's I tests on the residuals indicate that the main model results are not much affected by unobserved spatially correlated factors (cf. *SI Appendix*, section 1.7 and Table S3).

Model 6 extends model 5 (with $k = 13$) by including an interaction between class origin and neighborhood class (Fig. 5). Except for girls aged 17, we find no evidence that the social classes are influenced differently by the neighborhood class (*SI Appendix*, Tables S4 and S5). Hence, the high social class is not less affected by living in deprived neighborhoods, nor does this class benefit more from high-class neighbors, than individuals of lower-class origin. Moreover, at all ages, the children of high-class

SOCIAL SCIENCES

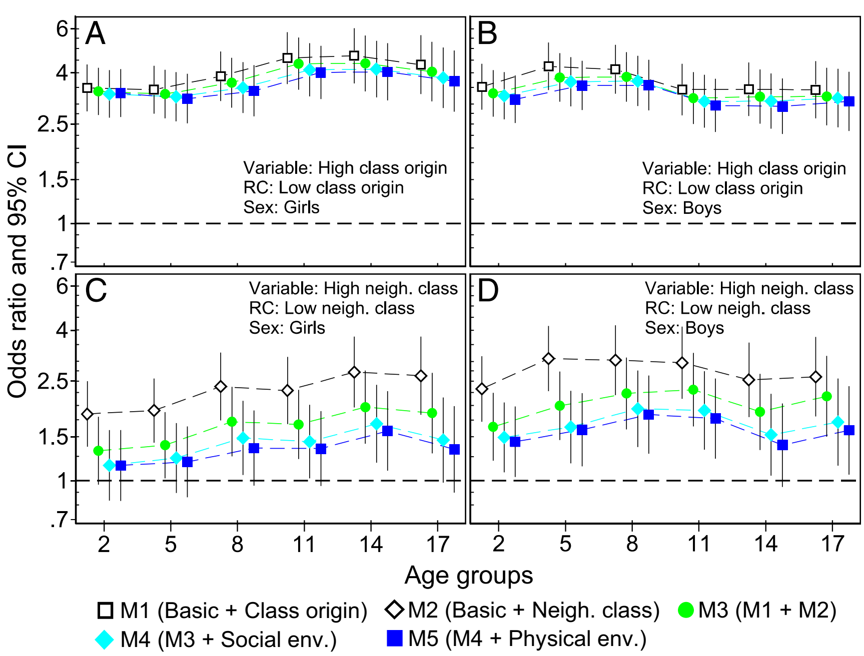


**Fig. 4.** Association between neighborhood conditions and other factors in childhood and having a university degree at age 40, Landskrona and Sweden, 1939 to 2015. Models 1 to 5, ages 2, 5, 8, 11, 14, and 17. Only the variables high-class origin and high neighborhood class are shown here (reference classes: low-class origin and low neighborhood class). (*A*) High-class origin, girls; (*B*) high-class origin, boys; (*C*) high neighborhood class, girls; and (*D*) high neighborhood class, boys. Girls: age 2, $n = 4{,}284$; age 5, $n = 4{,}117$; age 8, $n = 3{,}972$; age 11, $n = 3{,}960$; age 14, $n = 3{,}631$; age 17, $n = 3{,}369$). Boys: age 2, $n = 4{,}560$; age 5, $n = 4{,}316$, age 8, $n = 4{,}178$; age 11, $n = 4{,}057$; age 14, $n = 3{,}740$; and age 17, $n = 3{,}557$. The odds ratios and their 95% confidence intervals (bars) are plotted on a log scale.

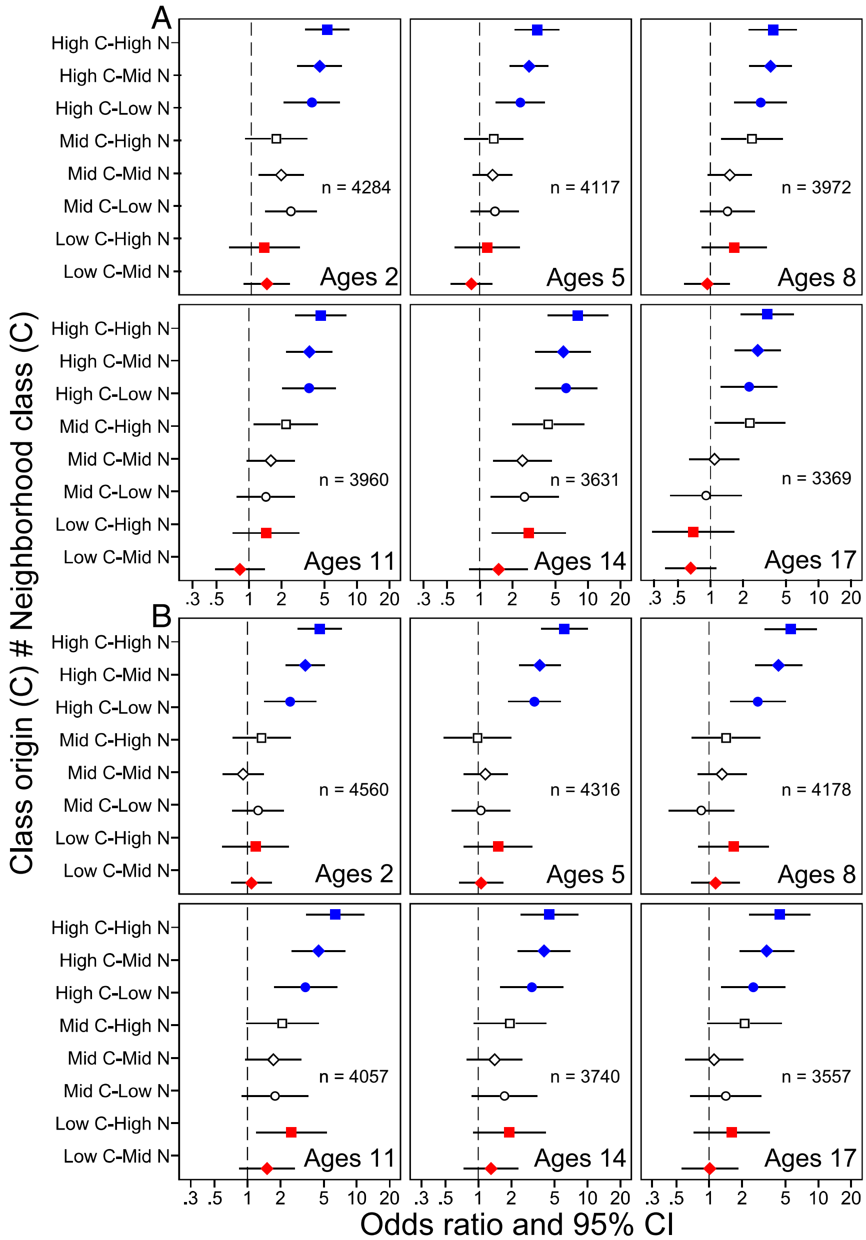


**Fig. 5.** Association between neighborhood conditions and other factors in childhood (ages 2, 5, 8, 11, 14, and 17) and having a university degree at age 40, Landskrona and Sweden, 1939 to 2015. Interactions are included between class origin and neighborhood class, using the control variables from model 5. Only the interaction variables are shown here (reference class: low-class origin # low neighborhood class). (*A*) Girls and (*B*) boys. The odds ratios and their 95% confidence intervals (bars) are plotted on a log scale. Note that different colors and symbol shapes are used to visually distinguish the class origin groups (high, mid, and low).

origin residing in a high-class neighborhood had the highest likelihood of all groups to have obtained a university degree by age 40. The highest odds are found for girls aged 14, among which those of high-class origin residing in a high-class neighborhood had 8.4 times higher odds of obtaining at a university degree than individuals of low-class origin residing in a low-class neighborhood.

Class origin alone also had a strong association with education: Children of high-class origin who resided in a low-class neighborhood still had a higher chance of obtaining a university degree than the mid- and low classes. However, at ages 14 for girls and 11 for boys (when the neighborhood influence was strong; Fig. 4 *C* and *D*), children of low-class origin residing in a high-class neighborhood had higher odds of obtaining a university degree than children of midclass origin residing in a low-class neighborhood.

## Discussion

We studied the association between neighborhood conditions throughout childhood and obtaining a university degree. By geocoding the longitudinal individual data at the address level, we were able to produce fine-scale and geographically weighted measures for a broad range ($k = 6, 13, 25, 50$, and 100) of the closest same-age neighbors throughout childhood. In addition, these measures captured aspects of the social and physical environment at the neighborhood level, possible peer effects at the smallest neighborhoods, and school districts. Because we did not rely on economic and demographic information from administrative units, we avoided some of the serious biases that may arise from the modifiable areal unit problem and the uncertain geographic context problem. Last, by having the child population of Landskrona linked to national registers, we avoided selection bias from only looking at the stayer population. Thus, these high-quality longitudinal data allowed us to contribute to the understanding of neighborhood influences at a very fine scale, during a period when higher education expanded greatly in Sweden.

Our findings demonstrate that even when growing up in a medium-sized town with relatively low economic inequality, the social class of the nearest (primarily 6, 13, and 25) neighbors of the same age throughout childhood was important for educational achievements, regardless of class origin and the elementary school children possibly attended. These findings are in line with other studies using larger samples but more aggregated data (e.g., refs. 17, 19, 21), as well as studies on peer effects (53, 54). Moreover, Chetty and Hendren (19) show that the longer children live in a neighborhood in which the residents have a higher income than their family has, the more their own income increases later in life. In relation to these findings, our results indicate that children from high-class families with a large share of high-class neighbors of the same age had consistently higher odds of obtaining a university degree than all other children. This relationship holds also when adjusting for the elementary school they most likely attended. The most striking example was found for the girls aged 14, among whom those of high-class origin in high-class neighborhoods had approximately eight times higher odds of obtaining a university degree than did those of low-class origin in low-class neighborhoods. Thus, not only did children growing up in low-class neighborhoods have lower educational attainment, but high-class children clustered with each other outperformed all other children. These findings indicate the importance of an advantageous background, both in terms of the immediate family and the neighborhood. Finally, even if all social classes benefited from the great expansion of higher education, the large inequalities between the classes remained unchanged throughout the period.

In smaller, more homogenous cities, segregation is usually more present at finer scales (36). Thus, using very few nearest neighbors to estimate neighborhood influences is more realistic for Landskrona and cities of similar types. The results from the models using different neighborhood sizes, in which the strongest associations and model fits were found for neighborhoods of 6, 13, and 25 children (*SI Appendix*, Figs. S3 and S4), support this claim. In addition, we modeled the influence of neighbors' class using a nonlinear Gaussian distance function. Such a function captures the dependency found in many spatial relationships better than linear distance functions or equal weights (51). Our sensitivity tests with exponential distance functions and equal weights indicate small differences in strength of associations and model fits between the weighting methods (*SI Appendix*, Figs. S5 and S6). This may be due to the small sizes of the neighborhoods, and we expect a larger impact of distance decay functions in larger neighborhoods. Therefore, by using fine-scale measures of individual neighborhoods and more realistic weighting methods, we are able to reveal patterns in areas that other measures and scales may not capture.

We cannot identify the precise mechanisms of the neighborhood influence, whether related to peer influence from children of similar age living nearby, or adults living or working in the neighborhood. By controlling for school districts, we capture some of the social capital obtained within the schools and that may influence children's academic success, such as after-school activities and bonds between teachers and other students (55). The fact that the neighborhood associations remain also when adjusting for school district, indicate an important role of the closer residential neighborhood. In addition, we find some indications of peer influence because the strongest neighborhood associations are found at the smallest scales. Hence, different mechanisms may be in play compared to those studies using much larger neighborhoods, such as Chetty et al. (17–20). Moreover, we expected that children from high-class origins would be less influenced by deprived neighborhoods than children of lower-class origins, as indicated by other studies (56, 57). One reason is that high-class children through their parents may have more protective buffer against external adversities (57), and that network ties are stronger within the same social classes (58). Our findings, in contrast, do not suggest that that the neighborhood influences in childhood differed depending on the class origin, and this finding did not change when adjusting for the elementary school they most likely attended. Moreover, the associations of neighborhood conditions and educational outcomes differed by both sex, age, and number of nearest neighbors. Girls were less influenced by their childhood neighbors, and between the ages 11 and 17, they were more influenced by their class origin, as opposed to boys. Mechanisms that may explain these results for the older children are different interaction patterns and differences in the size and power of the social networks across sexes and ages (59).

The limitations of our study relate to whether the conclusions can be applied to other urban settings as well. We do not claim that Landskrona is representative in a statistical sense, but there is no reason to assume that the influences of the social and physical environment would be very different for other similar places in Sweden and elsewhere. Moreover, although we did not identify any serious problems with spatial autocorrelation of the residuals, questions remain regarding how to adequately delimit the time intervals, which may have spatial correlations with unobserved factors. Last, we defined neighborhoods using snapshot information on yearly basis. Due to the nature of the data, however, it is possible to consider the full residential histories and estimate the cumulative neighborhood effects throughout childhood. Such index can be defined on, e.g., a monthly basis, or by updating it at each migration event within the city. The latter is computational intensive but may improve the predictions in future studies (12).

**ACKNOWLEDGMENTS.** This study is part of the research program "The Landskrona Population Study," funded by the Swedish Foundation for Humanities and Social Sciences (Riksbankens Jubileumsfond). It also forms part of the LONGPOP project, which has received funding from the European Union's Horizon 2020 research and innovation program under the Marie Skłodowska-Curie grant agreement 676060 (disclaimer: this publication reflects only the author's view and that the Agency is not responsible for any use that may be made of the information it contains). We thank the PNAS reviewers for very helpful comments, which improved the quality of this paper. We also thank Gabriel Brea Martinez for providing estimates of the Gini index of Landskrona.

1. S. E. Mayer, C. Jencks, Growing up in poor neighborhoods: How much does it matter? *Science* **243**, 1441–1445 (1989).
2. I. G. Ellen, M. A. Turner, Does neighborhood matter? Assessing recent evidence. *Hous. Policy Debate* **8**, 833–866 (1997).
3. M. L. Small, K. Newman, Urban poverty after the truly disadvantaged: The rediscovery of the family, the neighborhood, and culture. *Annu. Rev. Sociol.* **27**, 23–45 (2001).
4. P. Sharkey, J. W. Faber, Where, when, why, and for whom do residential contexts matter? Moving away from the dichotomous understanding of neighborhood effects. *Annu. Rev. Sociol.* **40**, 559–579 (2014).
5. J. W. Ainsworth, Why does it take a village? The mediation of neighborhood effects on educational achievement. *Soc. Forces* **81**, 117–152 (2002).
6. J. Brooks-Gunn, G. J. Duncan, P. K. Klebanov, N. Sealand, Do neighborhoods influence child and adolescent development? *Am. J. Sociol.* **99**, 353–395 (1993).
7. K. Burger, How does early childhood care and education affect cognitive development? An international review of the effects of early interventions for children from different social backgrounds. *Early Child. Res. Q.* **25**, 140–165 (2010).
8. B. Jensen, A. Seltzer, Neighbourhood and family effects in educational progress. *Aust. Econ. Rev.* **33**, 17–31 (2000).



SOCIAL SCIENCES

9. J. Crane, "Effects of neighborhoods on dropping out of school and teenage childbearing" in *The Urban Underclass*, C. Jencks, P. E. Peterson, Eds. (The Brookings Institution, 1991), pp. 299–320.
10. V. E. Lee, D. T. Burkam, *Inequality at the Starting Gate: Social Background Differences in Achievement as Children Begin School*, (ERIC, 2002).
11. R. Szulkin, J. O. Jonsson, *Ethnic Segregation and Educational Outcomes in Swedish Comprehensive Schools*, (Stockholm University, 2007).
12. G. T. Wodtke, D. J. Harding, F. Elwert, Neighborhood effects in temporal perspective: The impact of long-term exposure to concentrated disadvantage on high school graduation. *Am. Sociol. Rev.* **76**, 713–736 (2011).
13. L. Donnelly *et al*., Geography of intergenerational mobility and child development. *Proc. Natl. Acad. Sci. U.S.A.* **114**, 9320–9325 (2017).
14. E. K. Andersson, B. Malmberg, Contextual effects on educational attainment in individualised, scalable neighbourhoods: Differences across gender and social class. *Urban Stud.* **52**, 2117–2133 (2015).
15. C. Jencks, S. E. Mayer, The social consequences of growing up in a poor neighborhood. *Inner City Poverty United States* **111**, 111–186 (1990).
16. W. J. Wilson, *The Truly Disadvantaged: The Inner City, the Underclass, and Public Policy*, (Univ Chicago Press, Chicago, 2012).
17. R. Chetty, N. Hendren, L. F. Katz, The effects of exposure to better neighborhoods on children: New evidence from the moving to opportunity experiment. *Am. Econ. Rev.* **106**, 855–902 (2016).
18. R. Chetty, N. Hendren, P. Kline, E. Saez, Where is the land of opportunity? The geography of intergenerational mobility in the United States. *Q. J. Econ.* **129**, 1553–1623 (2014).
19. R. Chetty, N. Hendren, The impacts of neighborhoods on intergenerational mobility I: Childhood exposure effects. *Q. J. Econ.* **133**, 1107–1162 (2018).
20. R. Chetty, N. Hendren, The impacts of neighborhoods on intergenerational mobility II: County-level estimates. *Q. J. Econ.* **133**, 1163–1228 (2018).
21. R. Manduca, R. J. Sampson, Punishing and toxic neighborhood environments independently predict the intergenerational social mobility of black and white children. *Proc. Natl. Acad. Sci. U.S.A.* **116**, 7772–7777 (2019).
22. R. J. Sampson, S. W. Raudenbush, F. Earls, Neighborhoods and violent crime: A multilevel study of collective efficacy. *Science* **277**, 918–924 (1997).
23. R. J. Sampson, J. D. Morenoff, T. Gannon-Rowley, Assessing "neighborhood effects": Social processes and new directions in research. *Annu. Rev. Sociol.* **28**, 443–478 (2002).
24. C. Muller, R. J. Sampson, A. S. Winter, Environmental inequality: The social causes and consequences of lead exposure. *Annu. Rev. Sociol.* **44**, 263–282 (2018).
25. P. Sharkey, R. Sampson, "Neighborhood violence and cognitive functioning" in *Social Neuroscience: Brain, Mind, and Society*, R. Schutt, M. S. Keshavan, L. J. Seidman, Eds. (Harvard University Press, 2017).
26. A. L. Wani, A. Ara, J. A. Usmani, Lead toxicity: A review. *Interdiscip. Toxicol.* **8**, 55–64 (2015).
27. A. Clifford, L. Lang, R. Chen, K. J. Anstey, A. Seaton, Exposure to air pollution and cognitive functioning across the life course—a systematic literature review. *Environ. Res.* **147**, 383–398 (2016).
28. J. Feng, T. A. Glass, F. C. Curriero, W. F. Stewart, B. S. Schwartz, The built environment and obesity: A systematic review of the epidemiologic evidence. *Health Place* **16**, 175–190 (2010).
29. M. Van den Berg *et al*., Health benefits of green spaces in the living environment: A systematic review of epidemiological studies. *Urban Urban Green.* **14**, 806–816 (2015).
30. M. S. Granovetter, The strength of weak ties. *Am. J. Sociol.* **78**, 1360–1380 (1973).
31. M. S. Granovetter, The strength of weak ties: A network theory revisited. *Sociol. Theory* **1**, 201–233 (1983).
32. N. Lin, K. S. Cook, R. S. Burt, *Social Capital: Theory and Research*, (Transaction Publishers, 2001).
33. S. Issehnane, F. Sari, Neighborhood effects and peer effects what are the consequences for academic achievement? *Revue Économique* **64**, 775–804 (2013).
34. C. F. Manski, Identification of endogenous social effects: The reflection problem. *Rev. Econ. Stud.* **60**, 531–542 (1993).
35. G. Topa, Y. Zenou, "Neighborhood and network effects" in *Handbook of Regional and Urban Economics*, G. Duranton, J. V. Henderson, W. C. Strange, Eds. (Elsevier, 2015), pp. 561–624.
36. J. Östh, B. Malmberg, E. Andersson, "Analysing segregation using individualized neighbourhoods" in *Socio-Spatial Segregation: Concepts, Processes and Outcomes*, C. D. Lloyd, I. Shuttleworth, D. Wong, Eds. (Policy Press, 2014), pp. 135–162.
37. J. R. Logan, B. Bellman, Before the Philadelphia negro: Residential segregation in a nineteenth-century northern city. *Soc. Sci. Hist.* **40**, 683–706 (2016).
38. L. Hedman, D. Manley, M. Van Ham, J. Östh, Cumulative exposure to disadvantage and the intergenerational transmission of neighbourhood effects. *J. Econ. Geogr.* **15**, 195–215 (2015).
39. U. Türk, J. Östh, How much does geography contribute? Measuring inequality of opportunities using a bespoke neighbourhood approach. *J. Geogr. Syst.* **21**, 295–318 (2019).
40. S. Openshaw, Ecological fallacies and the analysis of areal census data. *Environ. Plann. A* **16**, 17–31 (1984).
41. A. S. Fotheringham, D. W. Wong, The modifiable areal unit problem in multivariate statistical analysis. *Environ. Plann. A* **23**, 1025–1044 (1991).
42. M. Kwan, The uncertain geographic context problem. *Ann. Assoc. Am. Geogr.* **102**, 958–968 (2012).
43. T. Bengtsson, M. Dribe, P. Svensson, L. Quaranta, The Scanian Economic Demographic Database (Machine-Readable Database) (Version 6.1, Centre for Economic Demography, Lund University, 2018).
44. Statistics Sweden, Mikrodata för Registret över totalbefolkningen (RTB). Registret över totalbefolkningen. https://www.scb.se/vara-tjanster/bestalla-mikrodata/vilka-mikrodata-finns/individregister/registret-over-totalbefolkningen-rtb/. Accessed 21 March 2020.
45. Statistics Sweden, Mikrodata för Registret över befolkningens utbildning. Utbildningsregistret (UREG). https://www.scb.se/vara-tjanster/bestalla-mikrodata/vilka-mikrodata-finns/individregister/registret-over-befolkningens-utbildning/. Accessed 21 March 2020.
46. L. Quaranta, STATA Programs for using the Intermediate Data Structure (IDS) to construct files for statistical analysis. *Hist. Life Course Stud.* **3**, 1–19 (2016).
47. M. Roderick, What's happening to the boys? Early high school experiences and school outcomes among African American male adolescents in Chicago. *Urban Educ.* **38**, 538–607 (2003).
48. M. Stanfors, *Education, Labor Force Participation and Changing Fertility Patterns. A Study of Women and Socioeconomic Change in Twentieth Century Sweden* (Lund Studies in Economic History, No. 22, Almqvist and Wiksell International, 2003).
49. M. H. Van Leeuwen, I. Maas, A. Miles, *HISCO: Historical International Standard Classification of Occupations*, (Leuven University Press, 2002).
50. M. H. Van Leeuwen, I. Maas, *HISCLASS: A Historical International Social Class Scheme. Universitaire Pers Leuven*, (Leuven University Press, 2011).
51. A. S. Fotheringham, C. Brunsdon, M. Charlton, *Geographically Weighted Regression: The Analysis of Spatially Varying Relationships*, (Wiley, 2003).
52. A. S. Fotheringham, C. Brunsdon, M. E. Charlton, "Geographically weighted regression" in *The Sage Handbook of Spatial Analysis*, A. S. Fotheringham, P. A. Rogerson, Eds. (Sage, 2008), pp. 243–254.
53. A. J. Mashburn, L. M. Justice, J. T. Downer, R. C. Pianta, Peer effects on children's language achievement during pre-kindergarten. *Child Dev.* **80**, 686–702 (2009).
54. G. J. Bobonis, F. Finan, Neighborhood peer effects in secondary school enrollment decisions. *Rev. Econ. Stat.* **91**, 695–716 (2009).
55. M. Virtanen, J. Ervasti, T. Oksanen, M. Kivimaki, J. Vahtera, "Social capital in schools" in *Global Perspectives on Social Capital and Health*, (Springer, 2013), pp. 65–85.
56. S. Pong, L. Hao, Neighborhood and school factors in the school performance of immigrants' children. *Int. Migr. Rev.* **41**, 206–241 (2007).
57. E. Greenman, K. Bodovski, K. Reed, Neighborhood characteristics, parental practices and children's math achievement in elementary school. *Soc. Sci. Res.* **40**, 1434–1444 (2011).
58. M. McPherson, L. Smith-Lovin, J. M. Cook, Birds of a feather: Homophily in social networks. *Annu. Rev. Sociol.* **27**, 415–444 (2001).
59. M. McPherson, L. Smith-Lovin, M. E. Brashears, Social isolation in America: Changes in core discussion networks over two decades. *Am. Sociol. Rev.* **71**, 353–375 (2006).