Inequalities in Educational Outcomes: How Important is the Family?*

Julia Bredtmann[1,2,3] and Nina Smith[1,4,5]

[1]*Aarhus University*
[2]*Ruhr University Bochum*
[3]*RWI Essen*
[4]*CESifo*
[5]*IZA Bonn*

This version: January 2015

Abstract

In this paper, we investigate sibling correlations in educational outcomes, which serve as a broad measure of the importance of family and community background. Making use of rich longitudinal survey and register data for Denmark, our main aim is to identify the parental background characteristics that are able to explain the resemblance in educational outcomes among siblings. We find sibling correlations in educational outcomes in the range of 15 to 33 percent, suggesting that up to a third of the variation in educational achievement can be explained by family and community background. Our results further reveal that parents' socio-economic background (i.e., their education, occupation, and income) can explain up to 44 percent of the sibling correlation. However, non-economic factors such as family structure, the incidence of social problems, and parents' educational preferences also play an important role for sibling similarities in educational outcomes.

JEL Classifications: I21, I24, J13

Keywords: Intergenerational mobility, sibling correlations, education

*The authors are grateful to seminar participants at the Ruhr University Bochum for helpful comments and suggestions. All remaining errors are our own. – This work was supported by a fellowship within the Postdoc-Program of the German Academic Exchange Service (DAAD). – All correspondence to: Julia Bredtmann, Department of Economics and Business, Aarhus University, 8210 Aarhus V, Denmark, Email: jbredtmann@econ.au.dk.

1 Introduction

There is a general interest in society to understand the importance of family background for individual achievement. In particular, social scientists have long been interested in exploring the intergenerational relationship between parents' and offspring's outcomes, such as their educational attainment or income. These studies are motivated by the aim to assess the degree of equality of opportunity in a society. Family background, broadly defined, represents circumstances that members of the offspring generation cannot be held accountable for, hence a strong dependence of individual outcomes on family background implies low equality of opportunity (cf. Roemer, 1998).

Though the concept of intergenerational mobility is certainly a meaningful one, the major limitation of traditional parent-offspring associations is that they are based on one single characteristic of the family. However, family background has an impact on children in many ways that cannot be picked up by one single variable.[1] An alternative approach to measure the importance of family background is to investigate the sibling correlation in economic outcomes. A sibling correlation can be interpreted as the fraction of the total variation in an outcome that can be attributed to factors shared by siblings. As such, sibling correlations provide a broad measure of the overall importance of family and community background.

Estimates of sibling correlations in educational outcomes, which are the focus of this paper, have centered around 0.5 to 0.6 for the US and 0.4 for Norway and Sweden. This suggests that even in the Nordic countries, which are characterized by an extensive welfare state and a long history of offering free post-secondary and higher education, 40 percent of the variation in educational outcomes can be attributed to family background. In our study, we focus on Denmark, a country that has been shown to rank at the top of the educational mobility scale.[2] Basically, we are interested in whether in a high-mobility country such as Denmark, inequalities in educational outcomes still exist.

Although sibling correlations give us an estimate of *how much* of the variation in educational outcomes can be attributed to family and community background, they do not tell us anything about *which* background characteristics matter for children's educational achievement. Björklund and Jäntti (2012) compare the sibling correlation in years of schooling with the respective intergenerational correlation between children's and parent's education and find that siblings share much more than their parents' education. Hence, if parental education is not of major importance, what exactly is it that makes siblings

[1] See Björklund and Jäntti (2012) for a nice and more extensive discussion of the limitations of traditional analysis of intergenerational mobility.

[2] In a cross-country comparison of the intergenerational correlation in years of schooling across 42 nations, among them 13 Western countries, Hertz *et al.* (2007) find Denmark to possess the highest level of intergenerational educational mobility among the Western countries, and one of the highest levels across the world.

similar in terms of their educational achievement?

One hypothesis is that because most siblings grow up in the same neighborhood, this could explain parts of the sibling similarity. However, recent studies for the US, the UK, and Sweden suggest that neighborhood characteristics are of minor relevance in explaining the sibling resemblance in educational outcomes. Hence, there must be something within the family that accounts for the relatively high sibling correlation in educational achievement. Obviously, parents influence their children via several channels beyond parental education: investments in their children's education, transmission of cultural values, attitudes, and social skills, and genetic endowments are all possible candidates. Moreover, in addition to the investment decisions and endowments of the parents, family members beyond the parental generation, such as grandparents, may influence the economic position of the child generation. From an equality-of-opportunity perspective, it is crucial to understand what it is that is so important about family background. A second aim of our paper is therefore to shed light on which family background characteristics are able to explain the sibling resemblance in educational outcomes.

Our contributions to the literature are manifold: First, we provide some first evidence on sibling similarities in educational outcomes for Denmark, adding upon previous literature for the US, the UK, and other Scandinavian countries. Second, we are the first to investigate whether family background is more important for obtaining an upper secondary educational degree or a tertiary educational degree, gaining insights into whether educational inequality increases or decreases at higher stages of the educational system. Lastly, we are the first to decompose the sibling correlation in educational outcomes in factors attributable to family and community characteristics, thereby considering a wide range of background characteristics, including parents' socio-economic status, cognitive skills and attitudes, as well as the role of grandparents and the neighborhood.

For a sample of children born between 1968 and 1984 in Denmark, we find sibling correlations in educational outcomes in the range of 15 to 33 percent, suggesting that up to a third of the variation in these outcomes can be explained by family and community background. For both brothers and sisters, family background is found to be more important for obtaining a tertiary educational degree than for obtaining an upper secondary degree, which suggests that educational inequality is higher at the top of the educational distribution. A decomposition of the sibling correlation reveals that parents' socio-economic status is the main determinant of sibling similarity in educational outcomes. However, non-economic factors such as family structure, the incidence of social problems, or parents' educational preferences also play an important role, especially in explaining sibling similarities in the completion of upper secondary education.

The outline of the papers is as follows. In Section 2, we summarize previous literature on the role of family background in economic outcomes and provide some basic informa-

tion on the educational system in Denmark. In Section 3, we explain our econometric approach to estimate and decompose the sibling correlations and describe the data and main variables used in our empirical analysis. The results of our analysis are presented in Section 4 and Section 5 concludes.

2 Background

2.1 Literature

A long literature in sociology and economics has aimed to estimate the importance of family background for children's future economic success. Most of this research has focused on the intergenerational relationship between parents' and offspring's outcomes.[3] Beginning with the study of Corcoran *et al.* (1976), researchers started to examine the sibling correlation as an alternative approach to measuring the importance of family background.

While intergenerational correlations in economic outcomes measure the relationship between parents' and offspring's economic success based on one single characteristic (e.g., years of education or earnings), sibling correlations in such outcomes provide a much broader measure of the role of family background for these outcomes. Measures of sibling similarity take into account not only the influence of the observed parental resource used in the intergenerational mobility analysis, but also all other unobserved factors that are shared by siblings and uncorrelated with the parental resource. Traditional studies of intergenerational associations in economic outcomes are therefore likely to study only "the tip of the iceberg" (Björklund and Jäntti, 2012, p. 471).

The majority of studies investigating sibling correlations in economic outcomes focus on investigations of sibling (or brother) correlations in permanent earnings or income. For the US, Solon *et al.* (1991), Levine and Mazumder (2007), and Mazumder (2008) find brother correlations in permanent earnings of about 0.45 to 0.50. Results for other countries as well as cross-country comparisons of sibling correlations in earnings (see, e.g., Björklund *et al.*, 2002; Schnitzlein, 2014) reveal that these estimates are of about the same size in Germany, while they are much lower in the Scandinavian countries. For Denmark, Schnitzlein (2014) estimates the sibling correlation in permanent earnings to be around 20 percent for both brothers and sisters, which is comparable with previous estimates for Finland (Österbacka, 2001; Björklund *et al.*, 2002), Norway (Björklund *et al.*, 2002) and Sweden (Björklund *et al.*, 2002, 2010; Björklund and Jäntti, 2012).

With respect to years of schooling and other educational outcomes, sibling correlations

[3]See Solon (1999) for a review of the earlier and Black and Devereux (2011) for a review of the more recent literature on intergenerational mobility.

are usually found to be higher than the respective correlations in income or earnings. For the US, Solon *et al.* (2000) and Mazumder (2008) estimate the sibling correlation in years of education to lie in the range of 0.5 to 0.6, suggesting that more than half of the variation in educational attainment in the US can be explained by family and community factors. Looking at test scores, Mazumder (2008, 2011) and Nicoletti and Rabe (2013) find similar results for the US and the UK, respectively. Again, economic inequality is lower in the Scandinavian countries: For Sweden, Björklund and Jäntti (2012) find an overall sibling correlation in years of schooling of 0.44, while the correlation is slightly higher for brothers (0.46) than for sisters (0.40). Raaum *et al.* (2006) and Lindahl (2011) obtain similar results using data for Norway and Sweden, respectively.

Only a few studies have tried to gain insights into which family and community factors drive the sibling correlation in economic outcomes. A part of the literature compares the sibling correlation in economic outcomes with the respective correlation in this outcome among neighboring children in order to impose a lower bound on the role of family background as opposed to neighbor and community effects for children's outcomes.[4] In general, these studies find a small role for neighborhoods in explaining sibling correlations in educational or economic outcomes. For instance, Lindahl (2011) finds brother and sister correlations in years of education of about 0.40, while the respective neighbor correlations are much smaller: 0.02 for males and 0.01 for females when basic family background characteristics (parental income and education) are accounted for.

Mazumder (2008) is the first to systematically decompose sibling correlations in economic outcomes into factors attributable to siblings' human capital (education, test scores), physical characteristics (height, weight, BMI), socially deviant behaviors (jail, drug use) and psychological characteristics (Rotter scale, self esteem). He finds that human capital can explain 50 percent or more of the brother correlation in wages and earnings, while non-cognitive measures such as deviant behavior and psychological characteristics can account for around 20 percent of these correlations. While Mazumder (2008) is mainly interested in identifying the underlying channels through which family and community affect children's future economic outcomes, Björklund *et al.* (2010) employ Mazumder's decomposition approach to investigate which specific characteristics of the parents are important for sibling similarities in long-run income. Using data on a sample of children born in 1953 who lived in the Stockholm metropolitan area in 1963, the authors find that parents' socio-economic status, as measured by parental education, income, as well as father's occupation, can only account for 13 percent (sisters) and 28 percent (brothers) of the raw sibling correlation in long-run income.[5] They further show that the explanatory

[4]See, amongst others, Solon *et al.* (2000) and Page and Solon (2003a,b) for the US, Nicoletti and Rabe (2013) for the UK, Raaum *et al.* (2006) for Norway, and Lindahl (2011) for Sweden.

[5]The raw sibling correlation in long-run income was estimated to be 0.23 for sisters and 0.25 for brothers.

power of the family characteristics rises to 58 percent for sisters and 71 percent for brothers when indicators of parents' involvement in schoolwork and parental attitudes are are added, suggesting that parental characteristics beyond parents' socio-economic status play a role for sibling similarities in long-run income.

In this paper, we contribute to the above literature in several ways: First, we add upon previous literature for the US and some European countries and provide first evidence on sibling similarities in educational outcomes for Denmark. In doing so, we go beyond the traditional analysis of years of schooling as an outcome variable, but explicitly investigate whether the role of family background varies over different stages of the educational system. Lastly, we are the first to apply a decomposition analysis as proposed by Mazumder (2008) to decompose the sibling correlation in educational outcomes in factors attributable to family and community characteristics. By making use of a combination of rich Danish survey and register data, we are able to consider a wide range of background characteristics, including parents' socio-economic status, cognitive skills and attitudes, as well as the role of grandparents and the neighborhood.

2.2 The Danish Educational System

[To be completed]

3 Method and Data

3.1 Method

The following statistical framework based on Solon *et al.* (1991) is used to measure the sibling correlation in educational outcomes. Each educational outcome (e.g., years of education) is denoted by y_{ij}, where j indexes siblings and i indexes families. The model for each outcome is then:

$$y_{ij} = \mu + \epsilon_{ij}, \qquad \epsilon_{ij} = a_i + b_{ij}, \tag{1}$$

where μ is the population mean and ϵ_{ij} is the residual. The latter can be decomposed into a permanent component common to all siblings in the family, a_i, and a permanent component that is individual-specific, b_{ij}, which captures individual deviations from the family component. Both a_i and b_{ij} are treated as random effects that are assumed to be independent of each other. The variance of y_{ij} is then simply:

$$\sigma^2_\epsilon = \sigma^2_a + \sigma^2_b. \tag{2}$$

The first term, σ_a^2, captures the variance in educational outcomes that is due to differences between families, whereas the second term, σ_b^2, captures the variance in educational outcomes within families. These two components are then used to calculate the correlation in permanent outcomes between siblings, ρ:

$$\rho = \frac{\sigma_a^2}{\sigma_a^2 + \sigma_b^2}. \tag{3}$$

This is also equivalent to the fraction of the overall variance in educational outcomes that is due to shared family and community background.

A sibling correlation can thus be thought of as an omnibus measure of the importance of family background and community effects. It includes the variance of anything shared by siblings, such as (observed and unobserved) parental resources and influences, as, e.g., parents' education and income, their parenting styles as well as their preferences and aspirations. Moreover, it captures things not directly related to the parents, such as school and neighborhood effects as well as the influence of other family members, as, e.g., the siblings' grandparents. However, there are also factors related to family and community background that are not captured by the sibling correlation, such as genetic traits not shared by siblings, differential treatment of siblings, and changes across time in the family, neighborhoods and schools. Therefore, the sibling correlation is a lower-bound measure of the importance of such factors.

Following, amongst others, Mazumder (2008, 2011), Björklund *et al.* (2010), and Lindahl (2011), the variance components that are needed to calculate the sibling correlation are estimated using restricted maximum likelihood (REML). REML has been shown to be superior to other estimation methods (as, e.g., ANOVA formulas) when the data are unbalanced, which is the case in our study because of varying family sizes. A drawback of using REML is that the error components a and b must be assumed to be normally distributed. For years of education this may be less problematic, but for our binary outcome variables the normality assumption may be more suspect. We therefore checked the robustness of our results by using ANOVA formulas instead of REML to calculate the error components, which did not change our results substantially.[6] The standard errors of the sibling correlations are calculated using the delta method.

To understand how different observable characteristics (e.g., parental education or income) influence the sibling correlation in educational outcomes, we follow the method developed by Mazumder (2008) and augment equation (1) with the vector X_{ij}, which contains different variables depending on the specification. These variables are treated as fixed effects in the REML framework and should reduce the residual variation in the

[6]This is in accordance with the results of Mazumder (2008, 2011), Lindahl (2011), and Nicoletti and Rabe (2013), who find that the results based on REML are similar to those of other estimation methods.

outcome variable. Hence, adding the control variables X_{ij} to the model should produce lower estimates of the family component (σ_a^{2*}) and the sibling correlation (ρ^*) than what was found without their inclusion. The relative difference between the two sibling correlations ($(\rho-\rho^*)/\rho$) can then be interpreted as an estimate of the fraction of the overall sibling correlation that can be attributed to the specific factors in question. This provides an upper-bound estimate of the causal effect because it includes all omitted factors that are also correlated with the included fixed effects. For example, the reduction in ρ due to the inclusion of parents' education would be comprised of both the direct effect of parents' on children's education as well as any omitted factors that are correlated with parents' education and influence children's educational outcomes (e.g., parents' cognitive skills or preferences). Implementing this approach for a wide variety of possible explanatory variables, either by including them one at a time or by including them simultaneously, should tell us something about which family background characteristics are critical to explaining the sibling correlation in educational outcomes.

Following Björklund *et al.* (2010), we start with adding basic individual and parental characteristics (i.e., parental education, occupation, and income) to our model to see how much standard measures of parental socio-economic background can add to explaining the sibling correlation in educational outcomes. Our main question, though, is which family characteristics beyond parents' socio-economic status are able to explain the sibling resemblance in educational outcomes. To answer this question, we then take the sibling correlation obtained from the model controlling for parents' socio-economic background, ρ^*, as our new baseline correlation and subsequently add new variables to the vector X_{ij}. The relative difference between this new sibling correlation ρ^{**} and ρ^* (i.e., $(\rho^* - \rho^{**})/\rho^*$) then gives us an estimate of the additional contribution of these family characteristics to the sibling correlation once parents' socio-economic status is already controlled for. Such an analysis is not only interesting in itself, but also reduces the problem of unobserved heterogeneity accruing from the fact that many family background characteristics (e.g., parents' cognitive skills) might be highly correlated with parent's socio-economic background. Therefore, they might mainly capture the indirect effect of parents' economic status on children's outcomes.[7] Nonetheless, we can not rule out that unobserved heterogeneity is still a problem in our analysis. The obtained estimates should therefore still be interpreted as upper-bound measures of the importance of the respective family characteristics for the sibling resemblance in educational outcomes.

[7]Our approach thus differs from Björklund *et al.* (2010) in that we explicitly look at the additional contribution of non-economic aspects of family background, while Björklund *et al.* (2010) are mostly interested in the overall contribution of economic and non-economic factors to the sibling correlation.

3.2 Data

Our basic data source is the Danish Longitudinal Survey of Youth (DLSY)[8], which is augmented with data from the Danish registers. The DLSY is an ongoing longitudinal study of a nationally representative sample of 3,151 Danish respondents who were born in or around 1954. The main respondents were first interviewed in 1968 when they were around 14 years old and attended 7th grade of elementary school.[9] The purpose of the 1968 DLSY survey was to analyze the determinants and consequences of educational achievement and attainment. For this purpose, the class teacher in the respondent's school class and one of the respondent's parents were also interviewed in 1968 and 1969, respectively. While the parents were asked about their own educational background and their preferences about education, the teacher had to give an assessment of the school class with respect to its proficiency level and its social structure. During the first interview, the DLSY respondents further took part in a 3-dimensional (verbal, spatial, inductive) intelligence test.

The main DLSY respondents have since been followed and interviewed in 1970, 1971, 1973, 1976, 1992, 2001, and finally in 2004 when they were around 50 years old. The follow-up surveys in 1970 and 1971 (age 16-17) and in 1973 and 1976 (age 19 and 22) tracked, among other things, respondents educational choices at the end of elementary school and secondary school, respectively, while the following surveys in 1992 (age 38), 2001 (age 47), and 2004 (age 50) provide rich information on issues such educational and occupational careers, family formation and fertility, attitudes and aspirations, health, social contacts and relationships, and leisure time activities.

Over the 36 years of data collection, the response rates in the DLSY have consistently remained very high. In the latest 2004 survey around 76 percent of the original 3,151 sample members were successfully interviewed. The drop in response rates over the period is partially due to respondents moving out of the country or dying. However, since the DLSY data can be linked with register data from Statistics Denmark, a considerable amount of information exists for all 3,151 original sample members even though they have dropped out of the DLSY survey.

The DLSY respondents constitute our index generation, which is the second generation included in the DLSY data. We then extend the survey data in several ways. By making use of parental identifiers in the Danish register data, we first identify all biological children of the DLSY respondents (born between 1968 and 2012), which constitute our sibling

[8]We use the "Cumulative 1968–2004 File" of the DLSY data. For a technical report of these data, see Jæger (2014).

[9]The first 1968 DLSY survey used a cluster based sampling. Among all the 7th grade elementary school classes in Denmark registered by the Ministry of Education in the school year 1967/68, a nationally representative sample of 152 (or just over 4 percent of all) school classes was selected and all pupils in these classes were included in the DLSY.

sample. In doing so, our sibling sample includes both multiples and singletons, the latter being useful for calculating the individual error component.[10] We then identify the second biological parent of these children, i.e., the (former) spouse of the DLSY respondent. Following most of the previous literature (e.g., Björklund *et al.*, 2009; Björklund and Jäntti, 2012; Mazumder, 2008), we then restrict our sample to siblings having the same biological mother and father. This leaves us with a raw sample of 5,281 children of 2,494 biological mothers and fathers.[11]

In principle, the parental identifiers in the Danish register data can also be used to identify the parents of the DLSY respondents (and their spouses), i.e., the first generation of individuals within the family. However, parental identifiers are only complete for children born in or after 1960, which leads to the fact that we are only able to identify a small proportion of grandparents in the register data (around ?%). While this is unfortunate, we can still use the register information for these individuals to improve the precision of our measures of grandparents' education in the survey data. Our resulting data set thus consists of three generations of individuals of the same families.

To measure children's educational success, we consider three different outcome variables obtained from the register of the level of education maintained by Statistics Denmark. This register provides a detailed code of the type of the highest completed education, the completion date and how many years of schooling the highest completed education corresponds to. The latter variable serves as our first outcome measure, which ranges from 9 years, the compulsory schooling grade, to 21 years for individuals holding a PhD degree. While completed years of education is an informative measure of an individual's overall educational success, we consider two further outcome variables. In order to be able to answer whether family background is more important at the bottom or at the top of the educational ladder, we define two binary indicators: (i) a variable that takes value 1 if the individual has completed upper secondary education and (ii) a variable that takes value 1 if the individual has completed tertiary education. All outcome measures are observed in 2012, the latest year of observation in the educational registers.

In looking at children's completed education, we have to restrict the sample to those individuals who are old enough to have completed their educational track at the end of our observation window. Hence, we restrict the sample to individuals aged 28 and older in 2012, the age at which the vast majority of individuals in Denmark has entered the labor market. This leaves us with a final sample of 3,087 children, born between 1968 and 1984, of 1,934 parents.

[10] By including singletons in the analysis, we follow Mazumder (2008, 2011). Mazumder (2008) also shows that including singletons has little effects on the estimated sibling correlations.

[11] It is important to note that while our index generation is nationally representative of all 7th grade pupils in 1968 (and therefore roughly representative of the 1954 birth cohort), our sample of children of these parents is not representative of the Danish population. The problem of a possible selectivity of our sample will be discussed in more detail later in this Section.

Our basic control variables at both the sibling and the parental level come from the Danish population and employment registers. In 1980, Denmark was the first country to conduct a totally register-based census, hence most variables are available from 1980 onwards (i.e., till 2012, the year of observation of our outcome variables). At the sibling level, we control for gender, age (and its square) as well as birth order, which has been shown to be highly relevant for children's educational outcomes. (e.g., Black *et al.*, 2005; Björklund and Jäntti, 2012)

At the parental level, our main control variables are measures of parents' socio-economic background, i.e., both parents' education, occupation, and income. Similarly to our outcome variable, parents' education is defined as mother's and father's completed years of education. Our measure of parents' occupation is a mixture of their labor force status and their occupational status over the period 1980 to 2012. For each year within this period, we observe parents' main economic status distinguishing between the self-employed, white-collar workers, blue-collar workers, other workers (not specified), the unemployed, and those out of the labor force. The predominant status over the 32 year period is then taken as our measure of parents' occupation.[12] Lastly, we control for both parents' logarithm of average income over the period 1980 to 2012.[13] Descriptive statistics of our outcome and basic control variables are shown in Table 1.

Regarding our outcome measures, we can see that at the age of 28, 86 percent of the individuals have completed upper secondary education and 41 percent have completed tertiary education. However, there are some differences between the genders, especially with respect to the completion of tertiary education. Women are 6 percentage point more likely to have obtained an upper secondary degree and even 21 percentage points more likely to have obtained a tertiary educational degree than men.[14] Considering individuals' completed years of education, which amounts to 14.3 years for men and 14.9 years for women, the difference between the brother and sister sample is less apparent. This is due to the fact that conditional on having obtained an upper secondary and a tertiary educational degree, respectively, women are more likely to have undertaken short or medium higher education, while men are relatively more likely to have undertaken long higher education and longer vocational education and training. The descriptive statistics at the sibling level further show that our sample is balanced with respect to the individuals' gender and that the mean age in our sample is 33 years for both brothers and sisters.

Regarding the parents' characteristics, the statistics show that both mothers and

[12]We also tried different definitions of our occupation variable. For instance, we controlled for mother's and father's share of years out of the labor market. This does not alter our results substantially.

[13]Income is defined as the sum of earned income, transfer income, property income (excl. imputed rent of owner-occupiers) and other non-classifiable income attributable directly to the individual, and before deduction of labor market contributions and special pension contributions. Income is measured in 2000 prices.

[14]The differences between the mean values for brothers and sisters is significantly different from zero.

fathers were less educated than their children and that men were more educated than women (13.1 years vs. 13.5 years), which is reversed in the next generation. The majority of individuals in the parents' generation are white-collar workers and blue-collar workers. Around 9 percent of the fathers are self-employed, while this value is much lower for mothers (around 2 percent). More than 10 percent of the women have been out of the labor force over most of their working life, while this value is smaller for men (6.6 percent). The share of long-term unemployed is relatively low among both genders (around 1.5 percent). The average income of men and women amounts to 283,226 DKK and 206,076 DKK, respectively, hence husbands' income exceeds wives' income by 37 percent.

One of our main concerns is that our sample of children may not be representative of the Danish population. In particular, the fact that our sibling sample consists of children born to members of the 1954 birth cohort creates a potential selectivity with respect to the parents' age at birth. By restricting our sample to children born before 1984, we have to exclude all children born to the initial DLSY respondent after the age of 30. This has two main consequences: First, our sample of children is likely to be negatively selected among all children of the respective birth cohort, as we observe children born to individuals who became parents relatively early in life. Second, first- and second-born children should be over-represented in our sample, while children of higher birth order are more likely to be excluded from our sample due to the described age restriction. Unfortunately, it is hard to tell how this potential selectivity may affect our estimation results.

While we are not able to determine the direction of bias in our sibling estimates, we can at least try to assess the extent of selectivity in our estimation sample. We do so by comparing our sibling sample to a representative sample of children that is not born to a specific birth cohort. In particular, we define a sample that is in the same age range as our sibling sibling sample, whereas the age range is defined by the median age in our sibling sample +/- 4 years.[15] For the resulting sample, which is a representative sample of the 1976 to 1984 birth cohort, we calculate the same descriptive statistics as in Table 1.

As can be seen from Table A1, the two samples are largely similar with respect to their observable characteristics. As expected, there is a small difference in the mean birth order between the two samples, which amounts to 1.78 in the representative sample as compared to 1.43 in our sibling sample. This sort of selectivity, however, does not translate into differences between siblings' educational outcomes An exception is the percentage of brothers holding a tertiary education degree, which is about 4 percentage points higher in the representative sample than in our sibling sample (0.35 vs. 0.31). Hence, our sample might be slightly negatively selected with respect to the tertiary educational attainment of men (though the difference in mean values between the samples is not statistically

[15]The median age in our sibling sample is 32 years, hence the sample consists of all individuals aged 28 to 36 in 2012.

different). Considering the characteristics of the siblings' parents, we also find hardly any differences between the two samples. This makes us confident that, though our sample is not representative of the Danish population, selectivity is not a major concern in our analysis.

The variables in Table 1 serve as our main control variables. One aim of this paper, however, is to investigate which family background characteristics beyond parents' socio-economic status are crucial for children's educational success and are thus able to explain the sibling resemblance in educational outcomes. In order to answer this question, we make use of the rich information on family background available in both the register and the survey data by defining a large set of further control variables which are potentially relevant for children's educational outcomes.

[To be completed]

4 Results

4.1 Raw Sibling Correlations in Educational Outcomes

The basic estimates of sibling correlations in our three educational outcomes, along with the respective estimated family and individual components, are reported in Table 2. Starting with the outcomes for mixed sexes, we find a sibling correlation of 0.33 for our years of education variable, suggesting that about a third of the variation in years of education can be explained by family and community background. This estimate is somewhat lower than previous estimates for Norway and Sweden, centering around 0.4 (cf. Raaum *et al.*, 2006; Lindahl, 2011; Björklund and Jäntti, 2012). This supports the finding of Hertz *et al.* (2007), who – based on intergenerational correlations in years of education – find Denmark to possess the highest level of intergenerational educational mobility among the Western countries, including Finland, Norway and Sweden.

Considering only same-sex siblings, we find a correlation in years of schooling of 31 percent for brothers and 39 percent for sisters, though the difference between the estimates is not statistically significant. This is in line with the results of Lindahl (2011) and Björklund and Jäntti (2012), who also find very small gender differences in sibling correlations in years of education for Sweden. Turning to our other outcome variables, we find a sibling correlation of 0.15 for the completion of upper secondary education and a correlation of 0.30 for the completion of tertiary education, suggesting that family background is twice as important for obtaining a higher educational degree than for obtaining a basic one. Both estimates are slightly higher for same-sex siblings, but of about a similar size for brothers and for sisters.

Summing up, we find sibling correlations in educational outcomes for Denmark that lie in the range of 0.15 to 0.39. Though these estimates are lower than comparable estimates for other Scandinavian countries and especially for the US (cf. Mazumder, 2008), they are still of a considerable magnitude. This is especially true if one considers that sibling correlations are lower-bound estimates of the overall importance of family and community background, because there are also factors attributable to the family that are not shared by siblings, such as differences in genes or a differential treatment of the siblings. While the above results are raw estimates that do not account for any individual or family background characteristics whatsoever, we now start to add control variables to our specification in order to explain which family characteristics make siblings similar in terms of their educational outcomes.

4.2 The Contribution of Parents' Socio-economic Status

Table 3 shows our estimates of the sibling correlations for mixed siblings when basic individual characteristics as well as indicators for mothers' and fathers' socio-economic background are (subsequently) added. For each of the three outcomes considered, we report both the estimated sibling correlation (column 1) and the percentage decrease in the sibling correlation due to adding the respective covariates (column 2). The latter statistic can be interpreted as an upper-bound measure of the contribution of the considered background characteristics to the sibling correlation in educational outcomes.

The first row displays the results when only some basic individual characteristics of the siblings (age, gender, and birth order) are controlled for. The estimated sibling correlations are somewhat smaller than the raw correlations shown in Table 2, but overall adding individual controls does not alter the results substantially. This is not surprising, given that controlling for individual characteristics should mainly sob up some of the residual variation previously captured by the individual component rather than by the family component.[16] In the next step, we separately add different indicators for parents' socio-economic status, starting with controls for mother's education, occupation, and income (rows 2 to 4). Considering siblings' years of education, we see that mother's education and occupation each can explain around 15 percent of the sibling correlation, while mother's income seems to be of less relevance (8 percent). The contribution of mother's background characteristics to the sibling correlation, however, varies over the stages of the educational system. It turns out that mother's occupation is able to explain a large part of the sibling correlation in completing upper secondary education (23 percent), while mother's education is relatively more important for obtaining a tertiary educational

[16] While this is true for gender and birth order, the siblings' age is certainly a factor related to family background, as – given that our siblings represent a sample of children born to members of a specific birth cohort – it partly captures the indirect effect of parents' age at birth.

degree (15 percent). A look at the full estimation results displayed in Table A2 reveals that with respect to mother's occupation, having an unemployed or non-participating mother seems to be most harmful for completing upper secondary education. Overall, mother's socio-economic status is able to explain 20 to 25 percent of the sibling similarity in educational outcomes (row 5).

Turning to the estimation results for father's socio-economic status (rows 6 to 9), we find similar results as for mother's socio-economic status. Among the background characteristics considered, father's education and occupation seem to be most relevant for explaining the sibling resemblance in educational outcomes, while the latter is again most important for obtaining an upper secondary degree. Father's income, on the other hand, is not able to explain more than 11 percent of the sibling correlation. Overall, father's socio-economic status can account for up to 27 percent of the sibling correlation in educational outcomes. Hence, mother's and father's socio-economic background seem to be about equally relevant for children's educational outcomes.[17]

In the last step, we add mother's and father's background characteristics simultaneously. First, we only include mother's and father's years of education (row 10), i.e., we look at the intergenerational correlation between parents' and children's education. Our results reveal that these factors alone are only able to explain between 16 and 24 percent of the sibling correlation in educational outcomes. Hence, there seems to be much more than parental education that is responsible for inequalities in educational outcomes. This supports our notion (and that of other researchers) that intergenerational relationships between parents' and children's outcomes are likely to only capture the "tip of the iceberg" (cf. Björklund and Jäntti, 2012).

In the last row of Table 3, we control for both mother's and father's socio-economic status. The results reveal that parental socio-economic background is able to explain about a third of the sibling correlation in years of education. The explanatory power of parents' background characteristics, however, is much higher for the completion of an upper secondary degree (44 percent) than for the completion of a tertiary degree (29 percent).

In Table 4, we have conducted the same analysis separately for brothers and sisters.[18] The results reveal that there are indeed some differences between the genders. With respect to the completion of upper secondary education, we find that mother's occupation and also her income are most relevant for explaining brother correlations in this outcome, while they are less relevant for sisters. On the other hand, father's occupation and income are more important for girls than for boys. For both genders, parents' education does

[17] Of course, these results have to be interpreted with some caution. Parents' education, occupation, and income are likely to be highly correlated with each other, making it difficult to ascertain the importance of one single indicator.

[18] Full estimation results are shown in Table A3.

hardly show any explanatory power for the sibling similarity in completing upper secondary education. Especially mother's years of education, however, are an important predictor of inequalities in the completion of tertiary education of girls. Overall, we find that parents' socio-economic status is more relevant for brothers regarding the completion of upper secondary education and more relevant for sisters regarding the completion of tertiary education.

The above results show that parental socio-economic status is a major determinant of inequalities in educational outcomes, especially of inequalities in obtaining a high school or vocational education degree. This result contradicts the findings of Björklund *et al.* (2010) for Sweden, who – focusing on income correlations – show that parental socio-economic background can explain no more than 28 percent of the sibling correlation in long-run income.[19] However, our results also reveal that the dominating part of the sibling correlations in these outcome cannot be explained by parents' education, occupation, or income. Hence, there must be something more than parents' socio-economic status that drives the sibling similarity in educational outcomes.

4.3 The Contribution of Other Family Characteristics

[To be completed]

4.4 The Contribution of the Neighborhood

In our search for factors that are able to explain the sibling similarity in educational outcomes, we have so far focused on background characteristics of the family, broadly defined. Another hypothesis would be that it is rather neighborhood characteristics shared by siblings than family characteristics that are able to explain inequalities in educational outcomes. In order to test this hypothesis, we follow previous literature (e.g., Solon *et al.*, 2000; Raaum *et al.*, 2006; Lindahl, 2011; Nicoletti and Rabe, 2013) and estimate correlations in educational attainment among children growing up in the same neighborhood in order to impose a lower bound on the role of family background in determining children's outcomes.

In order to measure the neighbor correlations in educational outcomes, we estimate a model similar to Eq. (1):

$$y_{cij} = \lambda + \nu_{cij}, \qquad \nu_{cij} = u_c + v_{cij}, \tag{4}$$

where y_{cij} is the educational outcome of individual j in family i in neighborhood c. λ is the population mean and ν_{cij} is the residual, which can be decomposed into a neighborhood

[19] We will discuss possible sources of differences between our results and the results of Björklund *et al.* (2010) in Section 4.3.

random component, u_c, and an individual-specific error term, v_{cij}. The variances of the individual and the neighborhood component can then be used to calculate the neighbor correlation in educational outcomes, $\varphi = \sigma_u^2/(\sigma_u^2 + \sigma_v^2)$, which captures the share of the between-neighborhood variation of the overall variance in educational outcomes.

As outlined by Solon *et al.* (2000), a neighbor correlation represents an upper-bound measure of the importance of the neighborhood for individuals' outcomes, as it captures both the "pure" neighborhood effect as well as an indirect effect, accruing from sorting of families into neighborhoods. As the sorting effect is assumed to be positive, the neighbor correlation as estimated from Eq. (4) represents an upper-bound measure capturing the indirect family effects as well.[20]

We define an individual's neighborhood in terms of where he or she lived at age 16. Though this might be an imperfect measure of the neighborhood, as the family may have lived elsewhere before, data limitations preclude us from following individuals further back in time. However, previous research has shown that even when families move, the neighborhoods to which they move are usually similar to the ones from where they move (Kunz *et al.*, 2003). We therefore assume that the neighborhood at age 16 is a relatively good proxy for the neighborhood environment the children grew up in.

We define neighborhoods at the level of the postcode area. There are 1,029 postcode areas in Denmark, which have an average population of 5,317 individuals.[21] Hence, postcode areas are quite broad and therefore probably not a perfect definition of a neighborhood. However, as discussed in previous literature, it is not clear what constitutes a good measure of the neighborhood: Small neighborhoods are likely to capture more of potential social interactions between members of the community, such as schoolmates and friends living in the same neighborhood ("peer effects"). On the other hand, one can imagine that individuals are influenced by the physical characteristics of their neighborhood, such as the general infrastructure, safety, and the like, even without direct social contacts with their neighbors.

In order to be able to compare the neighbor correlations with our previously estimated sibling correlations, we construct a neighborhood sample that resembles our sibling sample. Specifically, our neighborhood sample is comprised of all children born to the 1954 birth cohort. Moreover, we restrict the sample to individuals aged 28 to 44 in 2012, which

[20]Specifically, there are two different types of sorting, which are both assumed to positively contribute to the neighbor correlation: (i) sorting of similar families into the same neighborhood and (ii) sorting of advantaged families into advantaged neighborhoods. In order to tighten the bound of the neighbor correlation, some authors use measures of neighbor correlations that are adjusted for observed family characteristics. As our aim is to impose a lower bound on the role of the family in generating inequalities in educational outcomes, we estimate unadjusted neighbor correlations. For a formal derivation of the estimation of the neighborhood covariance, see Solon *et al.* (2000).

[21]The numbers refer to the year 2012. In 2007, Denmark underwent a local government reform, which involved a complete reorganization of municipalities and also lead to a change in postcodes. We therefore use the postcode areas as defined after 2007, even if observing individuals before 2007.

represent the minimum and maximum ages in our sibling sample.

The estimation results are shown in Table 7. Though the neighbor correlations are all statistically significant, they are very small in magnitude. The neighbor correlation in years of education, for instance, amounts to 0.024 for the mixed sample, while it is slightly higher for women than for men (0.026 vs. 0.022). Considering the other outcomes, the estimated neighbor correlations are even smaller. This suggests that in Denmark, less than three percent of the variation in educational outcomes can be explained by neighborhood effects. This result is in line with previous literature (e.g., Raaum *et al.*, 2006; Lindahl, 2011), which also find a small role for neighborhoods in explaining the sibling resemblance in educational outcomes.

Although the amount of variation explained by neighborhood correlations should exceed the explanatory power of standard regression-based neighborhood analysis (see, Page and Solon, 2003a), we further explore an alternative approach to measuring neighborhood effects, in which we add information on local neighborhood characteristics to our sibling estimates. First, we add information on population density, the average years of education, the unemployment rate, as well as the share of immigrants in the postcode area to our model. The results reveal that local neighborhood characteristics have hardly any explanatory power for sibling similarities in educational outcomes. As the choice of neighborhood variables is of course arbitrary, we next add postcode-area fixed effects to our sibling estimates. Again, the estimated sibling correlations in educational attainment remain largely unchanged. Lastly, we follow the argument of Raaum *et al.* (2006) that it might rather be regional characteristics than local conditions that affect children's outcomes and add municipality fixed effects to our sibling estimates. In this way, we should capture all of the variation in siblings' educational outcomes that is explained by the region in which they spent their childhood. Again, our estimated sibling correlations remain largely unchanged.[22] Hence, our conclusion that neighborhoods play a minor role in explaining sibling similarities in educational outcomes remains.

5 Conclusion

In this paper, we analyze the correlation in educational outcomes among siblings in Denmark. A sibling correlation captures everything that is shared by siblings and can thus be thought of as an omnibus measure of the importance of family and community background for individuals' outcomes. Our main contribution to previous literature is that we are not only interested in the extent of the sibling similarity in educational attainment, but further aim at identifying the determinants of educational inequalities across families.

For a sample of children born to participants in a Danish long-term study, which

[22] All estimation results are available from the authors upon request.

constitute a representative sample of all 7th graders in 1968, we find sibling correlations in educational outcomes in the range of 15 to 33 percent, suggesting that up to a third of the variation in these outcomes can be explained by family and community background. Hence, even in a highly egalitarian country such as Denmark, inequalities in educational attainment do exist.

For both brothers and sisters, we find family background to be more important for obtaining a tertiary educational degree than for obtaining an upper secondary degree, which suggests that educational inequality is higher at the top of the educational distribution. This result points to the possible existence of a "glass-ceiling effect" for children from disadvantaged families, which prevents them from entering upper levels of the educational system. If the general political aim is to reduce educational inequality, removing this barrier seems to be of particular importance.

A decomposition of the sibling correlation reveals that parents' socio-economic status is the main determinant of sibling similarities in educational outcomes. However, non-economic factors such as family structure, the incidence of social problems, or parents' educational preferences also play an important role, especially in explaining sibling similarities in the completion of upper secondary education. Parents' cognitive skills, their parenting styles, as well as neighborhood characteristics, in contrast, seem to be of minor relevance for explaining inequalities in educational outcomes. Our findings therefore reveal that the main determinants of the sibling similarity in educational outcomes are potentially susceptible to public policy, which suggests that there may indeed be scope for policy interventions to reduce inequalities in educational attainment.

Of course, there are also some limitations to our approach. First, we have to acknowledge that our results are purely descriptive and cannot be interpreted causally. In our view, it is a major challenge for future research to obtain causal inference on which family background characteristics are important for children's outcomes and to which extent these factors violate equality norms. Second, our analysis has been primarily empirical. As already noted by Björklund and Jäntti (2012), the search for factors that explain sibling similarities in economic outcomes should ideally be guided by an all-encompassing theoretical model. While the standard Becker-Tomes model (Becker and Tomes, 1979, 1986) provides a theoretical model of the causal effect of parental income on offspring's outcomes, the literature on sibling correlations would clearly benefit from a broader theoretical framework that explains how family and community factors interact to influence children's outcomes.

References

Becker, G. S. and Tomes, N. (1979). An Equilibrium Theory of the Distribution of Income and Intergenerational Mobility. *Journal of Political Economy*, **87** (6), 1153–1189.

— and — (1986). Human Capital and the Rise and Fall of Families. *Journal of Labor Economics*, **4** (3), 1–39.

Björklund, A., Eriksson, T., Jäntti, M., Raaum, O. and Österbacka, E. (2002). Brother correlations in earnings in Denmark, Finland, Norway and Sweden compared to the United States. *Journal of Population Economics*, **15** (4), 757–772.

— and Jäntti, M. (2012). How important is family background for labor-economic outcomes? *Labour Economics*, **19** (4), 465–474.

—, Jäntti, M. and Lindquist, M. J. (2009). Family background and income during the rise of the welfare state: Brother correlations in income for Swedish men born 1932–1968. *Journal of Public Economics*, **93** (5–6), 671–680.

—, Lindahl, L. and Lindquist, M. J. (2010). What More Than Parental Income, Education and Occupation? An Exploration of What Swedish Siblings Get from Their Parents. *The B.E. Journal of Economic Analysis & Policy*, **10** (1), Article 102.

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

—, — and Salvanes, K. G. (2005). The More the Merrier? The Effect of Family Size and Birth Order on Children's Education. *The Quarterly Journal of Economics*, **120** (2), 669–700.

Corcoran, M., Jencks, C. and Olneck, M. (1976). The Effects of Family Background on Earnings. *The American Economic Review Papers and Proceedings*, **66** (2), 430–435.

Hertz, T., Jayasundera, T., Piraino, P., Selcuk, S., Smith, N. and Verashchagina, A. (2007). The Inheritance of Educational Inequality: International Comparisons and Fifty-Year Trends. *The B.E. Journal of Economic Analysis & Policy*, **7** (2), Article 10.

Jæger, M. M. (2014). Danish Longitudinal Survey of Youth (DLSY) – Cumulative 1968–2004 File. The Danish National Centre for Social Research Working Paper ??:2014.

Kunz, J., Page, M. E. and Solon, G. (2003). Are point-in-time measures of neighborhood characteristics useful proxies for children's long-run neighborhood environment? *Economics Letters*, **79** (2), 231–237.

Levine, D. I. and Mazumder, B. (2007). The Growing Importance of Family: Evidence from Brothers' Earnings. *Industrial Relations*, **46** (1), 7–21.

Lindahl, L. (2011). A comparison of family and neighborhood effects on grades, test scores, educational attainment and income – evidence from Sweden. *The Journal of Economic Inequality*, **9** (2), 207–226.

Mazumder, B. (2008). Sibling similarities and economic inequality in the US. *Journal of Population Economics*, **21** (3), 685–701.

— (2011). Family and community influences on health and socioeconomic status: sibling correlations over the life course. *The B.E. Journal of Economic Analysis & Policy*, **11** (3), Article 1.

Nicoletti, C. and Rabe, B. (2013). Inequality in Pupils' Test Scores: How Much do Family, Sibling Type and Neighbourhood Matter? *Economica*, **80** (318), 197–218.

Österbacka, E. (2001). Family Background and Economic Status in Finland. *Scandinavian Journal of Economics*, **103** (3), 467–484.

Page, M. E. and Solon, G. (2003a). Correlations between Brothers and Neighboring Boys in Their Adult Earnings: The Importance of Being Urban. *Journal of Labor Economics*, **21** (4), 831–855.

— and — (2003b). Correlations between sisters and neighbouring girls in their subsequent income as adults. *Journal of Applied Econometrics*, **18** (5), 545–562.

Raaum, O., Salvanes, K. G. and Sørensen, E. Ø. (2006). The Neighbourhood is Not What it Used to be. *The Economic Journal*, **116** (508), 200–222.

Roemer, J. E. (1998). *Equality of opportunity*. Cambridge, MA: Harvard University Press.

Schnitzlein, D. D. (2014). How important is the family? Evidence from sibling correlations in permanent earnings in the USA, Germany, and Denmark. *Journal of Population Economics*, **27** (1), 69–89.

Solon, G. (1999). Intergenerational mobility in the labor market. In O. Ashenfelter and D. Card (eds.), *Handbook of Labor Economics, Vol. 3*, Amsterdam: Elsevier, pp. 1761–1800.

—, Corcoran, M., Gordon, R. and Laren, D. (1991). A Longitudinal Analysis of Sibling Correlations in Economic Status. *The Journal of Human Resources*, **26** (3), 509–534.

—, Page, M. E. and Duncan, G. J. (2000). Correlations between Neighboring Children in Their Subsequent Educational Attainment. *The Review of Economics and Statistics*, **82** (3), 383–392.

Tables

Table 1: Descriptive Statistics

	All siblings		Brothers		Sisters	
	Mean	StdD	Mean	StdD	Mean	StdD
Years of education	14.619	(2.311)	14.292	(2.284)	14.942	(2.293)
Completed upp. sec. education	0.860	(0.347)	0.832	(0.374)	0.887	(0.316)
Completed tertiary education	0.413	(0.492)	0.308	(0.462)	0.516	(0.500)
Female	0.503	(0.500)	–	–	–	–
Age in 2012	32.965	(3.261)	32.976	(3.256)	32.955	(3.267)
Birth order	1.430	(0.587)	1.427	(0.581)	1.432	(0.593)
Observations	3,087		1,534		1,553	

	Fathers		Mothers	
	Mean	StdD	Mean	StdD
Years of education	13.525	(2.822)	13.055	(2.646)
Main occupation, 1980-2012				
Self-employed	0.093	(0.290)	0.022	(0.146)
White-collar worker	0.402	(0.490)	0.507	(0.500)
Blue-collar worker	0.392	(0.488)	0.296	(0.457)
Other worker (not specified)	0.033	(0.178)	0.058	(0.235)
Unemployed	0.015	(0.122)	0.016	(0.124)
Out of the labor force	0.066	(0.248)	0.102	(0.303)
Log of average income, 1980-2012	12.554	(0.492)	12.236	(0.374)
Observations	1,934		1,934	

Table 2: Raw Sibling Correlations in Educational Outcomes

	All siblings	Brothers	Sisters
Years of education			
Sibling correlation	0.327	0.309	0.392
StdE	(0.026)	(0.058)	(0.046)
Family component	1.750	1.618	2.061
StdE	(0.046)	(0.099)	(0.066)
Individual component	3.602	3.621	3.199
StdE	(0.020)	(0.042)	(0.038)
Observations	3,087	1,534	1,553
Completed upp. sec. education			
Sibling correlation	0.151	0.213	0.188
StdE	(0.028)	(0.059)	(0.056)
Family component	0.018	0.030	0.019
StdE	(0.096)	(0.143)	(0.153)
Individual component	0.102	0.110	0.081
StdE	(0.020)	(0.040)	(0.037)
Observations	3,087	1,534	1,553
Completed tertiary education			
Sibling correlation	0.299	0.341	0.344
StdE	(0.027)	(0.053)	(0.048)
Family component	0.073	0.073	0.086
StdE	(0.050)	(0.084)	(0.076)
Individual component	0.170	0.141	0.164
StdE	(0.020)	(0.041)	(0.037)
Observations	3,087	1,534	1,553

Notes: – Estimates are produced using restricted maximum likelihood (REML). – The standard errors of the sibling correlations are calculated by using the delta method.

Table 3: Sibling Correlations in Educational Outcomes and Parents' Socioeconomic Background

	Years of education		**Completed upp. sec. ed.**		**Completed tertiary ed.**	
	ρ	% ↓	ρ	% ↓	ρ	% ↓
Individual controls						
Sibling correlation	0.285		0.133		0.256	
StdE	(0.028)		(0.028)		(0.028)	
Mother's education						
Sibling correlation	0.244	14.503	0.121	9.160	0.217	15.235
StdE	(0.028)		(0.028)		(0.028)	
Mother's occupation						
Sibling correlation	0.243	14.624	0.103	22.844	0.219	14.204
StdE	(0.028)		(0.028)		(0.028)	
Mother's income						
Sibling correlation	0.263	7.823	0.122	8.363	0.239	6.611
StdE	(0.028)		(0.028)		(0.028)	
Mother's education, occupation, and income						
Sibling correlation	0.224	21.339	0.099	25.158	0.203	20.437
StdE	(0.028)		(0.028)		(0.028)	
Father's education						
Sibling correlation	0.247	13.224	0.119	10.141	0.233	8.983
StdE	(0.028)		(0.028)		(0.028)	
Father's occupation						
Sibling correlation	0.250	12.103	0.106	20.590	0.230	9.903
StdE	(0.028)		(0.028)		(0.028)	
Father's income						
Sibling correlation	0.263	7.588	0.118	10.867	0.243	4.818
StdE	(0.028)		(0.028)		(0.028)	
Father's education, occupation, and income						
Sibling correlation	0.223	21.809	0.097	27.086	0.215	16.107
StdE	(0.028)		(0.028)		(0.028)	
Parents' education						
Sibling correlation	0.218	23.501	0.112	15.903	0.203	20.478
StdE	(0.028)		(0.028)		(0.028)	
Parents' education, occupation, and income						
Sibling correlation	0.185	34.931	0.075	43.568	0.182	28.666
StdE	(0.029)		(0.028)		(0.028)	

Notes: – Estimates are produced using restricted maximum likelihood (REML). – The standard errors of the sibling correlations are calculated by using the delta method. – The individual controls (age, gender, birth order) are included in all regressions and the results displayed in the first row constitute our new baseline estimates. – Full estimation results are shown in Table A2.

Table 4: Brother and Sister Correlations in Educational Outcomes and Parents' Socioeconomic Background

	Years of education				Completed upp. sec. education				Completed tertiary education			
	Brothers		Sisters		Brothers		Sisters		Brothers		Sisters	
	ρ	% ↓	ρ	% ↓	ρ	% ↓	ρ	% ↓	ρ	% ↓	ρ	% ↓
Individual controls												
Sibling correlation	0.270		0.350		0.214		0.160		0.290		0.299	
StdE	(0.059)		(0.049)		(0.059)		(0.057)		(0.057)		(0.050)	
Mother's education												
Sibling correlation	0.259	4.085	0.305	12.712	0.214	0.137	0.155	2.734	0.273	5.809	0.254	14.968
StdE	(0.057)		(0.050)		(0.057)		(0.057)		(0.057)		(0.051)	
Mother's occupation												
Sibling correlation	0.244	9.613	0.312	10.696	0.177	17.593	0.152	4.850	0.269	7.317	0.270	9.559
StdE	(0.059)		(0.050)		(0.060)		(0.056)		(0.057)		(0.051)	
Mother's income												
Sibling correlation	0.241	10.573	0.329	6.004	0.192	10.333	0.156	2.468	0.278	4.207	0.280	6.154
StdE	(0.061)		(0.050)		(0.060)		(0.057)		(0.058)		(0.051)	
Mother's education, occupation, and income												
Sibling correlation	0.238	11.645	0.289	17.399	0.179	16.698	0.152	5.003	0.263	9.313	0.247	17.407
StdE	(0.058)		(0.051)		(0.060)		(0.056)		(0.057)		(0.052)	
Father's education												
Sibling correlation	0.253	6.060	0.320	8.402	0.202	5.572	0.160	−0.154	0.283	2.584	0.279	6.731
StdE	(0.058)		(0.050)		(0.059)		(0.056)		(0.056)		(0.051)	
Father's occupation												
Sibling correlation	0.253	6.099	0.308	12.004	0.195	9.223	0.129	19.248	0.281	3.331	0.265	11.316
StdE	(0.059)		(0.051)		(0.059)		(0.057)		(0.057)		(0.052)	
Father's income												
Sibling correlation	0.270	−0.016	0.316	9.609	0.210	2.091	0.145	9.297	0.292	−0.510	0.273	8.547
StdE	(0.058)		(0.050)		(0.058)		(0.057)		(0.057)		(0.051)	
Father's education, occupation, and income												
Sibling correlation	0.240	10.878	0.283	19.145	0.185	13.671	0.128	19.884	0.274	5.573	0.247	17.167
StdE	(0.058)		(0.052)		(0.059)		(0.057)		(0.056)		(0.052)	
Parents' education												
Sibling correlation	0.244	9.650	0.290	17.235	0.203	5.376	0.156	2.111	0.267	8.005	0.245	17.947
StdE	(0.056)		(0.050)		(0.058)		(0.056)		(0.056)		(0.051)	
Parents' education, occupation, and income												
Sibling correlation	0.212	21.494	0.254	27.420	0.153	28.541	0.124	22.205	0.250	13.775	0.223	25.283
StdE	(0.059)		(0.052)		(0.061)		(0.056)		(0.057)		(0.052)	

Notes: – Estimates are produced using restricted maximum likelihood (REML). – The standard errors of the sibling correlations are calculated by using the delta method. – The individual controls (age, gender, birth order) are included in all regressions and the results displayed in the first row constitute our new baseline estimates. – Full estimation results are shown in Table A3.

Table 5: Contribution of Family Characteristics to Sibling Correlations

	Years of education		Completed upp. sec. ed.		Completed tertiary ed.	
	ρ	% ↓	ρ	% ↓	ρ	% ↓
Family structure						
Sibling correlation	0.163	11.854	0.053	27.418	0.177	3.843
StdE	(0.029)		(0.028)		(0.029)	
Social problems						
Sibling correlation	0.184	5.618	0.076	26.292	0.183	0.074
StdE	(0.030)		(0.030)		(0.030)	
Parent's cognitive skills						
Sibling correlation	0.179	2.664	0.071	2.019	0.181	2.265
StdE	(0.029)		(0.029)		(0.029)	
Parent's school efforts						
Sibling correlation	0.185	0.717	0.091	4.618	0.194	−0.182
StdE	(0.030)		(0.030)		(0.030)	
Parent's educational preferences						
Sibling correlation	0.217	4.517	0.109	14.127	0.206	0.678
StdE	(0.033)		(0.033)		(0.032)	
Parent's work preferences						
Sibling correlation	0.212	1.184	0.146	−1.708	0.203	2.884
StdE	(0.032)		(0.032)		(0.032)	
Parent's forward looking behavior						
Sibling correlation	0.215	2.274	0.134	2.056	0.191	4.474
StdE	(0.032)		(0.032)		(0.032)	
Parent's self-confidence						
Sibling correlation	0.202	0.376	0.096	−0.922	0.186	0.705
StdE	(0.031)		(0.031)		(0.031)	
Parent's parenting style						
Sibling correlation	0.181	3.008	0.154	−0.864	0.164	4.693
StdE	(0.034)		(0.034)		(0.034)	
Parent's cultural capital						
Sibling correlation	0.187	−0.573	0.127	0.805	0.168	−2.288
StdE	(0.032)		(0.032)		(0.032)	
Grandparents' soecio-economic background						
Sibling correlation	0.183	1.490	0.070	6.446	0.183	−0.135
StdE	(0.029)		(0.028)		(0.029)	
Grandparents' educational preferences						
Sibling correlation	0.186	−0.582	0.077	−2.488	0.183	−0.554
StdE	(0.029)		(0.028)		(0.028)	
Parent's school class FE						
Sibling correlation	0.173	4.921	0.074	−6.949	0.167	6.906
StdE	(0.030)		(0.030)		(0.030)	

Notes: – Estimates are produced using restricted maximum likelihood (REML). – The standard errors of the sibling correlations are calculated by using the delta method. – In all models, the sibling and parental characteristics included in Table 3 are additionally controlled for.

Table 6: Overall contribution of Family Characteristics

	Sibling correlation		Contribution to sibling correlation (in %)			
	raw	all controls	sibling char.[a]	parents' SES[b]	family char.[c]	all controls[d]
Years of education						
Sibling correlation	0.314	0.166	16.843	28.255	11.564	47.238
StdE	(0.037)	(0.042)				
Completed upp. sec. education						
Sibling correlation	0.212	0.103	13.942	18.757	30.713	51.557
StdE	(0.038)	(0.042)				
Completed tertiary education						
Sibling correlation	0.271	0.165	11.605	23.877	9.722	39.253
StdE	(0.038)	(0.042)				

Notes: – Estimates are produced using restricted maximum likelihood (REML). – The standard errors of the sibling correlations are calculated by using the delta method. – [a] Sibling characteristics entail all individual characteristics controlled for in Table 3. – [b] Parents' SES refers to all parental characteristics controlled for in Table 3. – [c] Family characteristics refers to controls in Table 5. – [d] All controls entails all control variables in Table 3 and 5.

Table 7: Raw Neighbor Correlations

	Both sexes	Men	Women
Years of education			
Neighborhood correlation	0.024	0.022	0.026
StdE	(0.002)	(0.003)	(0.003)
Observations	70,273	35,684	34,589
Completed upp. sec. education			
Neighborhood correlation	0.014	0.016	0.014
StdE	(0.002)	(0.002)	(0.002)
Observations	70,273	35,684	34,589
Completed tertiary education			
Neighborhood correlation	0.020	0.019	0.020
StdE	(0.002)	(0.002)	(0.002)
Observations	70,273	35,684	34,589

Notes: – Estimates are produced using restricted maximum likelihood (REML). – The standard errors of the sibling correlations are calculated by using the delta method. – The neighborhood is comprised of all individuals living in the same postcode area at age 16.

Appendix

Table A1: Descriptive Statistics, whole Danish population aged 28 to 36

	All siblings		**Brothers**		**Sisters**	
	Mean	StdD	Mean	StdD	Mean	StdD
Years of education	14.649	(2.412)	14.392	(2.416)	14.915	(2.379)
Completed upp. sec. education	0.851	(0.356)	0.822	(0.383)	0.880	(0.324)
Completed tertiary education	0.432	(0.495)	0.350	(0.477)	0.516	(0.500)
Female	0.492	(0.500)	–	–	–	–
Age in 2012	32.220	(2.583)	32.220	(2.585)	32.220	(2.582)
Birth order	1.783	(0.901)	1.784	(0.900)	1.783	(0.902)
Observations	460,899		234,324		226,575	

	Fathers		**Mothers**	
	Mean	StdD	Mean	StdD
Years of education	13.203	(3.085)	12.958	(2.858)
Main occupation, 1980-2012				
Self-employed	0.117	(0.321)	0.029	(0.169)
White-collar worker	0.382	(0.486)	0.455	(0.498)
Blue-collar worker	0.354	(0.478)	0.274	(0.446)
Other worker (not specified)	0.031	(0.173)	0.060	(0.238)
Unemployed	0.017	(0.129)	0.019	(0.136)
Out of the labor force	0.099	(0.299)	0.163	(0.369)
Log of average income, 1980-2012	12.536	(0.484)	12.158	(0.486)
Observations	346,214		346,214	

Table A2: Siblings' Educational Outcomes and Parents' Socioeconomic Background

	Years of education			Completed upp. sec. education			Completed tertiary education		
	Coef/StdE	Coef/StdE	Coef/StdE	Coef/StdE	Coef/StdE	Coef/StdE	Coef/StdE	Coef/StdE	Coef/StdE
Siblings									
Female	$0.618^{\dagger}$	$0.612^{\dagger}$	$0.594^{\dagger}$	$0.055^{\dagger}$	$0.054^{\dagger}$	$0.050^{\dagger}$	$0.203^{\dagger}$	$0.202^{\dagger}$	$0.198^{\dagger}$
	(0.078)	(0.075)	(0.074)	(0.012)	(0.012)	(0.012)	(0.016)	(0.016)	(0.016)
Age in 2012	0.621^{***}	0.610^{***}	0.552^{***}	0.070^{**}	0.068^{**}	0.053	0.087^{*}	0.084^{*}	0.077^{*}
	(0.220)	(0.212)	(0.211)	(0.035)	(0.034)	(0.034)	(0.047)	(0.045)	(0.045)
Age in 2012 (squared)	$-0.011^{\dagger}$	-0.010^{***}	-0.009^{***}	-0.001^{**}	-0.001^{**}	-0.001^{*}	-0.002^{**}	-0.001^{**}	-0.001^{*}
	(0.003)	(0.003)	(0.003)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Birth order	$-0.442^{\dagger}$	-0.175^{**}	−0.097	$-0.041^{\dagger}$	-0.019^{*}	−0.011	$-0.082^{\dagger}$	-0.030^{**}	−0.014
	(0.070)	(0.069)	(0.069)	(0.011)	(0.011)	(0.011)	(0.015)	(0.015)	(0.015)
Mother									
Years of education	–	$0.193^{\dagger}$	$0.129^{\dagger}$	–	$0.015^{\dagger}$	0.009^{***}	–	$0.041^{\dagger}$	$0.026^{\dagger}$
		(0.017)	(0.019)		(0.003)	(0.003)		(0.004)	(0.004)
Main occupation, 1980-2012 (Ref.: White-collar worker)									
Self-employed	–	–	−0.161	–	–	−0.014	–	–	−0.097
			(0.291)			(0.045)			(0.062)
Blue-collar worker	–	–	-0.319^{***}	–	–	−0.012	–	–	$-0.109^{\dagger}$
			(0.106)			(0.016)			(0.023)
Other worker (not specified)	–	–	-0.373^{**}	–	–	−0.033	–	–	$-0.134^{\dagger}$
			(0.186)			(0.028)			(0.040)
Unemployed	–	–	$-1.304^{\dagger}$	–	–	$-0.215^{\dagger}$	–	–	-0.149^{**}
			(0.337)			(0.052)			(0.072)
Out of the labor force	–	–	$-0.627^{\dagger}$	–	–	$-0.120^{\dagger}$	–	–	-0.101^{***}
			(0.172)			(0.026)			(0.037)
Log of average income, 1980-2012	–	–	0.214	–	–	0.012	–	–	0.033
			(0.135)			(0.021)			(0.029)
Father									
Years of education	–	$0.156^{\dagger}$	$0.111^{\dagger}$	–	$0.014^{\dagger}$	$0.009^{\dagger}$	–	$0.026^{\dagger}$	$0.017^{\dagger}$
		(0.016)	(0.016)		(0.002)	(0.003)		(0.003)	(0.003)
Main occupation, 1980-2012 (Ref.: White-collar worker)									
Self-employed	–	–	0.159	–	–	0.060^{***}	–	–	−0.022
			(0.151)			(0.023)			(0.032)
Blue-collar worker	–	–	-0.266^{***}	–	–	0.000	–	–	$-0.090^{\dagger}$
			(0.101)			(0.015)			(0.021)
Other worker (not specified)	–	–	-0.467^{**}	–	–	−0.044	–	–	-0.086^{*}
			(0.237)			(0.036)			(0.050)
Unemployed	–	–	-0.980^{**}	–	–	$-0.235^{\dagger}$	–	–	−0.095
			(0.383)			(0.060)			(0.082)
Out of the labor force	–	–	-0.489^{**}	–	–	-0.092^{***}	–	–	−0.065
			(0.193)			(0.030)			(0.041)
Log of average income, 1980-2012	–	–	$0.360^{\dagger}$	–	–	0.033^{**}	–	–	$0.065^{\dagger}$
			(0.092)			(0.014)			(0.020)
Constant	6.807^{*}	0.633	−4.224	−0.045	−0.545	−0.762	−0.532	-1.719^{**}	-2.490^{***}
	(3.683)	(3.567)	(4.100)	(0.581)	(0.577)	(0.654)	(0.780)	(0.758)	(0.873)
Sibling correlation	0.285	0.218	0.185	0.133	0.112	0.075	0.256	0.203	0.182
StdE	(0.028)	(0.028)	(0.029)	(0.028)	(0.028)	(0.028)	(0.028)	(0.028)	(0.028)
Observations	3,087	3,087	3,087	3,087	3,087	3,087	3,087	3,087	3,087

Notes: – † $p < 0.001$; *** $p < 0.01$; ** $p < 0.05$; * $p < 0.1$. – *Estimates are produced using restricted maximum likelihood (REML).* – *The standard errors of the sibling correlations are calculated by using the delta method.*

Table A3: Brothers' and Sisters' Educational Outcomes and Parents' Socioeconomic Background

	Years of education		Completed upp. sec. education		Completed tertiary education	
	Brothers Coef/StdE	Sisters Coef/StdE	Brothers Coef/StdE	Sisters Coef/StdE	Brothers Coef/StdE	Sisters Coef/StdE
Siblings						
Age in 2012	0.886^{***}	0.188	0.109^{**}	−0.004	0.118^{*}	0.040
	(0.305)	(0.295)	(0.052)	(0.044)	(0.062)	(0.066)
Age in 2012 (squared)	-0.013^{***}	−0.003	-0.002^{**}	−0.000	-0.002^{**}	−0.001
	(0.005)	(0.004)	(0.001)	(0.001)	(0.001)	(0.001)
Birth order	−0.084	−0.113	−0.018	−0.008	−0.006	−0.018
	(0.101)	(0.095)	(0.017)	(0.014)	(0.020)	(0.021)
Mother						
Years of education	$0.112^{\dagger}$	$0.142^{\dagger}$	0.009^{**}	0.010^{**}	$0.019^{\dagger}$	$0.032^{\dagger}$
	(0.026)	(0.026)	(0.004)	(0.004)	(0.005)	(0.006)
Main occupation, 1980-2012 (Ref.: White-collar worker)						
Self-employed	−0.642	0.271	−0.035	0.027	-0.156^{*}	−0.049
	(0.406)	(0.402)	(0.068)	(0.058)	(0.083)	(0.089)
Blue-collar worker	-0.248^{*}	-0.413^{***}	0.001	−0.027	$-0.128^{\dagger}$	-0.092^{***}
	(0.148)	(0.145)	(0.025)	(0.021)	(0.030)	(0.032)
Other worker (not specified)	−0.421	−0.368	−0.000	-0.065^{*}	$-0.206^{\dagger}$	−0.067
	(0.260)	(0.256)	(0.044)	(0.037)	(0.053)	(0.057)
Unemployed	$-1.525^{\dagger}$	-1.174^{**}	$-0.263^{\dagger}$	-0.163^{**}	-0.182^{**}	−0.130
	(0.445)	(0.482)	(0.075)	(0.070)	(0.091)	(0.107)
Out of the labor force	-0.704^{***}	-0.595^{**}	-0.124^{***}	$-0.114^{\dagger}$	-0.134^{***}	−0.075
	(0.235)	(0.238)	(0.039)	(0.034)	(0.048)	(0.052)
Log of average income, 1980-2012	0.134	0.243	0.026	−0.003	0.007	0.056
	(0.195)	(0.178)	(0.033)	(0.026)	(0.040)	(0.039)
Father						
Years of education	$0.119^{\dagger}$	$0.097^{\dagger}$	0.012^{***}	0.006^{*}	$0.020^{\dagger}$	0.014^{***}
	(0.023)	(0.023)	(0.004)	(0.003)	(0.005)	(0.005)
Main occupation, 1980-2012 (Ref.: White-collar worker)						
Self-employed	0.176	0.083	0.070^{*}	0.049^{*}	−0.024	−0.032
	(0.220)	(0.199)	(0.037)	(0.029)	(0.045)	(0.044)
Blue-collar worker	−0.216	-0.285^{**}	−0.010	0.012	-0.089^{***}	-0.083^{***}
	(0.138)	(0.140)	(0.023)	(0.020)	(0.028)	(0.031)
Other worker (not specified)	-0.844^{**}	−0.156	-0.093^{*}	−0.006	-0.158^{**}	−0.013
	(0.330)	(0.327)	(0.055)	(0.047)	(0.068)	(0.072)
Unemployed	−0.454	-1.416^{***}	−0.148	$-0.322^{\dagger}$	−0.011	−0.153
	(0.543)	(0.513)	(0.092)	(0.075)	(0.111)	(0.114)
Out of the labor force	-0.593^{**}	−0.275	-0.132^{***}	−0.047	−0.048	−0.069
	(0.262)	(0.280)	(0.044)	(0.040)	(0.054)	(0.062)
Log of average income, 1980-2012	0.294^{***}	$0.626^{\dagger}$	0.036^{*}	0.037	0.049^{**}	0.121^{***}
	(0.112)	(0.168)	(0.019)	(0.024)	(0.023)	(0.037)
Constant	−7.997	−1.097	-1.968^{**}	0.452	-2.636^{**}	-2.627^{**}
	(5.855)	(5.798)	(0.996)	(0.860)	(1.191)	(1.290)
Sibling correlation	0.212	0.254	0.153	0.124	0.250	0.223
StdE	(0.059)	(0.052)	(0.061)	(0.056)	(0.057)	(0.052)
Observations	1,534	1,553	1,534	1,553	1,534	1,553

Notes: – † $p < 0.001$; *** $p < 0.01$; ** $p < 0.05$; * $p < 0.1$. – *Estimates are produced using restricted maximum likelihood (REML).* – *The standard errors of the sibling correlations are calculated by using the delta method.*