



Published in final edited form as:

B E J Econom Anal Policy. 2011 September 6; 11(3): . doi:10.2202/1935-1682.2876.

Family and Community Influences on Health and Socioeconomic Status: Sibling Correlations Over the Life Course*

Bhashkar Mazumder*

*Federal Reserve Bank of Chicago, bmazumder@frbchi.org

Abstract

This paper presents new estimates of sibling correlations in health and socioeconomic outcomes over the life course in the U.S. Sibling correlations provide an omnibus measure of the importance of all family and community influences. I find that sibling correlations in a range of health and socioeconomic outcomes start quite high at birth and remain high over the life course. The sibling correlation in birth weight is estimated to be 0.5. Sibling correlations in test scores during childhood are as high as 0.6. Sibling correlations in adult men's wages are also around 0.5. Decompositions provide suggestive evidence on which pathways may account for the gradients in health and SES by family background. For example, sibling correlations in cognitive skills and non-cognitive skills during childhood are lower controlling for family income. Similarly, parent education levels can account for a sizable portion of the correlation in adult health status among brothers.

Keywords

sibling correlations; health; SES

I. Introduction

There is a growing recognition among economists that health is a critical component of human capital and that it plays a crucial role in determining socioeconomic status (Currie, 2009). Recent work has emphasized that key periods of child development may be especially susceptible to environmental conditions that affect health capital which in turn may affect socioeconomic success (Cunha and Heckman, 2007). There is also a long-standing literature in the social sciences that has documented gradients in adult health and mortality by socioeconomic status (SES), suggesting that there are important effects running in the other direction, i.e. from SES to health (e.g. Grossman, 2005). Therefore it would be valuable to gain greater insight into the patterns in the relationship between health and SES and how they arise over the lifecycle.

Since there are a large number of potential variables that can be used for such an analysis, it is not obvious how to best summarize lifecycle patterns. A potentially useful way to reduce the dimensionality of the issue is to focus on the extent to which inequality in health and SES is due to differences in family background. This paper uses such an approach by estimating sibling correlations in measures of health and SES over various stages of the life course in the U.S. utilizing variance decomposition models on the Panel Study of Income

*I gratefully acknowledge a small grants award from the PSID. I thank the two referees for their helpful comments. I also thank participants at the conference on "SES and Health across Generations and over the Life Course" for their comments.

Dynamics (PSID) data. Sibling correlations provide a straightforward summary measure of the combined effects of family background and community influences on a particular outcome. The measure estimates the fraction of the overall variation that is attributable to differences across families and therefore provides a simple gauge of inequality in SES and health due to family background.

Although this study focuses primarily on documenting sibling correlations in a wide variety of measures, the variance decomposition approach also provides a simple and useful framework for including covariates (Mazumder, 2008). Therefore, one can examine the extent to which particular family background characteristics (e.g. family income, parent education) can account for the sibling correlation in each outcome. This may potentially provide some “first order” evidence to guide future research that uses other research designs to uncover causal estimates.

The main finding is that sibling correlations in several key measures of health and socioeconomic outcomes start quite high at birth and remain high over the life course. For example, the sibling correlation in birth weight is estimated to be 0.5. Sibling correlations in test scores during childhood are as high as 0.6 and sibling correlations in adult men’s wages are also around 0.5. While previous research using sibling correlations has shown that family background plays an important role in explaining long-term economic success in the U.S. (e.g. Solon et al 1991, Mazumder, 2008), the findings in this study further suggest that the effects of family background may be quite large even early in life and that they are also evident for health outcomes.

Decompositions that use a set of limited covariates also suggest that for certain outcomes, the inclusion of variables such as family income can reduce the sibling correlation. For example, a portion of the inequality (across families) in children’s cognitive and non-cognitive skills and future adult male earnings can be accounted for by differences in family income. Future research should consider utilizing this approach with a wider variety of covariates in order to investigate the effects of other potential aspects of family environment. An important caveat, however, is that these are purely descriptive results and cannot be interpreted causally. Therefore, research must also supplement such analysis with studies that utilize research designs that are more appropriate for inferring causal effects. Nevertheless, the findings in this paper at least raise the possibility that there may be scope for policy interventions to reduce inequality in these outcomes.

The rest of the paper proceeds as follows. Section II provides some background material and a brief discussion of some of the prior literature. Section III presents the methodology. Section IV describes the PSID data in greater detail. Section V presents the results and Section VI concludes.

II. Background and Previous Literature

A long literature in sociology and economics has tried to estimate the importance of family background on socioeconomic outcomes. Early studies were often missing variables on key family background characteristics which led to the concern that family background might matter more than what the available measures would indicate. As a result, researchers began to examine the sibling correlation as an alternative approach to measuring the importance of family background (e.g. Corcoran et al 1976). Conceptually, the sibling correlation provides a summary statistic that captures *all* of the effects of sharing a common family as well as any other shared factors (e.g. common neighborhoods, school quality).¹ If the similarity in say,

¹Conversely, many aspects of family background including genetic traits and sibling-specific parental behaviors will not be captured.

general health status between siblings, is not much different compared to randomly chosen individuals, then we would expect a small correlation. If, however, a large fraction of the variance in health is due to factors common to growing up in the same family environment then the correlation might be sizable.

For brevity, I briefly discuss a few studies of the sibling correlation in economic outcomes.² Several previous studies have exploited the panel dimension of the PSID to estimate sibling correlations in economic outcomes in the U.S. using multiple years of data and have generally estimated correlations of around 0.3 to 0.4 (e.g. Solon et al 1991; Bjorklund et al 2002; Page and Solon, 2003). Other U.S. studies have utilized the NLS original cohort of young men and found broadly similar results (Altonji and Dunn, 1991; Ashenfelter and Zimmerman, 1997; Levine and Mazumder, 2007). Levine and Mazumder (2007) and Mazumder (2008) found larger estimates of around 0.5 when utilizing the NLSY79. A few studies outside of the U.S. have also estimated sibling correlation in economic outcomes, generally finding much lower correlations in Nordic countries (Bjorklund et al 2002; Raaum et al 2006; Bjorklund et al 2009; Bjorklund et al 2010). In a recent working paper, Schnitzlein (2011) finds that compared to the U.S., the sibling correlation is lower in Denmark but nearly the same in Germany.

A few studies have decomposed sibling correlations. Altonji and Dunn (2000) use a factor model to decompose sibling correlations in various labor market outcomes. They find that unobserved preferences play an important role in explaining sibling correlations in labor supply. Solon et al (2000) and Page and Solon (2003) decompose sibling correlations in schooling and earnings into factors that may be related to neighborhood effects and generally interpret their findings to suggest a small role for neighborhoods. Raaum et al (2006) and Lindahl (forthcoming) also find a small role for neighborhoods in explaining sibling correlations in educational or economic outcomes in Nordic countries. Mazumder (2008) decomposes sibling correlations in economic outcomes into factors attributable to human capital (education, test scores), physical characteristics (height, weight, BMI), socially deviant behaviors (jail, drug use) and psychological characteristics (Rotter scale, self esteem). Mazumder finds that human capital can explain 50 percent or more of the brother correlation in wages and earnings. He also finds that non-cognitive measures such as deviant behavior and psychological characteristics can also account for around 20 percent of these correlations. However, Mazumder does not utilize health as an outcome or as a covariate for SES outcomes and he does not examine sibling correlations over the life course.

Following Mazumder, Bjorklund et al (2010) employ the same decomposition approach to estimate which specific characteristics of parents (in addition to parent income) may account for the sibling correlation in income using Swedish data. They find that incorporating parent attitudes and parental involvement can account for some portion of the observed sibling correlation.

III. Methodology

I utilize the following statistical framework.³ Each outcome is denoted by y_{ijt} , where i indexes families, j indexes siblings and t indexes years. Although the majority of outcomes

²There are many studies of sibling correlations in non-economic outcomes both in the U.S. and in other countries which I do not discuss here.

³The notation here follows Solon et al (1991) and has also been used by Bjorklund et al (2002) and Mazumder (2008) among others.

considered here do not vary by year, the notation is kept general to encompass those cases.⁴ Outcomes are then modeled as follows:

$$y_{ijt} = \beta X_{ijt} + \varepsilon_{ijt} \quad (1)$$

The vector X_{ijt} will typically contain age dummies to account for lifecycle effects and a female dummy when both sexes are pooled. For economic outcomes, year effects are included to account for business cycle conditions. The residual, ε_{ijt} , which is purged of these effects, is then decomposed as follows:

$$\varepsilon_{ijt} = a_i + u_{ij} + v_{ijt} \quad (2)$$

The three terms on the right hand side of (2) are treated as random effects that are independent of each other by construction.⁵ The first term, a_i , is the permanent component that is common to all siblings in family i . The second term, u_{ij} , is the permanent component that is individual-specific. For outcomes for which I use repeated observations on the same individual, v_{ijt} represents the transitory component that reflects noise due to either temporary shocks or measurement error in the survey.⁶ In the case of outcomes that are not repeated, v_{ijt} is omitted. The variance of, for example, age-adjusted earnings, ε_{ijt} , then is simply:

$$\sigma_{\varepsilon}^2 = \sigma_a^2 + \sigma_u^2 + \sigma_v^2 \quad (3)$$

The first term, σ_a^2 , captures the variance in the permanent component that is due to differences *between* families while the second term, σ_u^2 , captures the variance in the permanent component that is due to differences *within* families. These two components are then used to construct the correlation in permanent outcomes between siblings, ρ .

$$\rho = \frac{\sigma_a^2}{\sigma_a^2 + \sigma_u^2}, \quad (4)$$

Some earlier studies such as Solon et al (1991) and Bjorklund et al (2002) used a two step approach to estimate the variance components in this so called “mixed model” (mixed because it contains both fixed effects, or regressors, and random effects). First they use a regression to estimate (1) and to produce the residuals and they use analysis of variance (ANOVA) formulas on the residuals. Due to the unknown properties of ANOVA when the data are unbalanced (e.g. differing number of siblings per formula), Mazumder (2008) used Restricted Maximum Likelihood (REML). REML has a number of advantages such as consistency, asymptotic normality, and a known asymptotic sampling dispersion matrix. However, REML requires an assumption that the data are normally distributed. For some of the outcomes to be considered here (e.g. log wages, birth weight), this is not likely a major factor. For other outcomes such as health status on a 1 to 5 scale (which is treated as continuous variable) or indicator variables, the assumption of normality is not appropriate and, therefore, those results should be treated more speculatively. The appropriate statistical model that can handle cases of ordered or dichotomous outcomes is not used in this paper

⁴The time varying outcomes include test scores where there are multiple scores from tests taken at different ages, and the various long-run economic outcomes.

⁵Conceptually, this framework allows one to divide the permanent component into a part that is perfectly correlated among siblings and a part that is perfectly uncorrelated among siblings. Previous studies have found that there is little or no *cross-sectional* correlation in the transitory component.

⁶The model can also be extended to account for serial correlation in the transitory component.

but would be a useful contribution for future work. An alternative approach utilized by Altonji and Dunn (1991), Solon et al (2000), Page and Solon (2003) and Bjorklund et al (2009) is to use method of moments. One downside is that judgment calls are required in how exactly to weight families of different sizes. For simplicity I only use REML in this analysis. Standard errors are calculated using the delta method.

To decompose the sibling correlation, I add a few additional regressors (race, family income, parent education and parent health) to the vector X in (1). Their inclusion may reduce the residual variation in the outcome variable and may alter the share of the residual variance due to the family and individual components, potentially leading to a smaller sibling correlation.

Previous economic studies have taken advantage of repeated observations on individuals to remove the effects of v_{ijt} on the estimated correlation. In principle the same approach can be taken for *non-economic* outcomes. For example, it is well known that test scores can be a quite noisy measure of underlying knowledge or skill. Therefore, multiple observations on test scores can be used to produce error-corrected measures of the sibling correlation.

IV. Data

The Panel Study of Income Dynamics began in 1968 with a representative sample of over 18,000 individuals living in 5,000 families. These families have subsequently been re-interviewed each year through 1997 and thereafter biennially. All persons in the original PSID families are followed in subsequent years and anyone born to or adopted by PSID sample members are themselves followed. When children leave their parents' homes, they are classified as a new PSID family and are then interviewed in each wave. The PSID sample therefore includes numerous groups of siblings who have been tracked for as long as 40 years.

This analysis uses two distinct groups of cohorts in the PSID to cover sibling correlations over the lifecycle. I use the 1985 to 1997 birth cohorts that were included in the Child Development Supplement (CDS) for estimating sibling correlations in birth outcomes, childhood outcomes and adolescent outcomes.⁷ I use cohorts born between 1951 and 1968 for measuring the sibling correlations in adult outcomes. The CDS sample is restricted to include only those who were the son or daughter of the household head in 1997. Similarly the adult sample is restricted to those who were the son or daughter of the household head in 1968. The adult sample is further restricted to those who become household heads or wives of a household head. The adult sample is confined to include only those who were members of the nationally representative portion of the sample. Similar to most of the previous literature, I do not restrict either sample to include only biological siblings since the question of interest concerns family background broadly defined. Both samples also include singletons which are useful for calculating the individual component.

The full sample size for the CDS cohorts meeting these restrictions is 3246 individuals in 2177 families. The actual sample sizes, however, will vary depending on the outcome being measured. For some CDS outcomes for which I make use of repeated observations from the same individuals, the actual number of observations used for the estimation will be considerably higher. The full adult sample includes 3265 individuals from 1355 families.

⁷See the User Guide for CDS-III for more detailed information including background on the the tests described below. (http://psidonline.isr.umich.edu/CDS/cdsiii_userGd.pdf)

For the CDS sample, I begin the analysis with a set of birth outcomes which includes: birth weight⁸ (converted to grams), gestation length (measured in days), health at birth (3 point scale going from better than average to worse than average), poor health at birth (an indicator for worse than average health at birth) and NICU (an indicator if a newborn was admitted to a neonatal intensive care unit). The sample means and other summary statistics for all the samples are shown in Appendix Table A1.

I next use the CDS sample to measure childhood health. A useful broad measure of childhood health is health status rated on a 1 to 5 scale (Excellent, Very Good, Good, Fair, Poor) by the parent or care giver. This is measured at three different age ranges (0 to 5, 6 to 10, and 11 to 22). Other measures include whether the child was reported to have had any health limitation, learning disability, speech impairment, emotional problem, allergy, anemia, asthma, developmental delays, diabetes, hyperactivity, and height and weight. Although the CDS asks questions concerning an even larger set of health outcomes, the incidence rates were far too low to be meaningful for many of these outcomes (e.g. autism). For some of the outcomes I restricted the age ranges so that the incidence rates are meaningful. I also include age dummies in all of the analysis.

The next stage of analysis examines a range of childhood educational measures that reflect both resources and outcomes. These include whether the parent expects the child to go to college and an indicator for whether the child has fewer than 10 books. I then examine four specific tests scores from the Woodcock-Johnson tests of achievement: letter-word identification, passage comprehension, calculation and applied problems. I estimate sibling correlations at two age ranges, 6 to 10 and 11 to 15 as well as pooling the two age ranges. All of the measures are first converted to z-scores by age and year of test. Finally, I make use of the WISC Digit Span Test for short term memory.

I also utilize the Transition to Adulthood (TA) modules given to older cohorts of the CDS in 2005 and 2007 to assess sibling similarities in measures that are of particular relevance for adolescents. These include whether students finished high school (does not include GED) and whether they ever enrolled in a college. I limit those outcomes to those at least 18 years of age. I also examine high school grade point average (GPA).⁹ The TA survey includes an extremely rich and detailed set of questions that assess adolescent attitudes and behaviors. I use two composite measures that summarize answers to questions concerning mental and emotional wellbeing. Both of these measures provide repeated observations so I pool all responses and calculate the sibling correlation in the “permanent” component. Finally, I include a set of measures dealing with other relevant health and addictive behaviors: whether an individual ever smoked cigarettes, used diet pills, used amphetamines or used marijuana.

The final set of outcomes utilizes siblings drawn from the original 1968 households. The economic outcomes include log annual earnings, log wages, log annual hours worked and log family income. Additionally I examine years of completed education, health status and disability.

V. Results

Birth Outcomes

The results for birth outcomes are shown in Table 1. For birth weight I find a sizable sibling correlation of 0.53 that does not change much if I limit the sample to only boys or only girls.

⁸Following some previous studies I do not include the birth weight of individuals who were born prematurely.

⁹I also experimented with SAT math and reading scores but the samples with valid scores were small.

The standard error is quite small at 0.02. This is similar to a sibling correlation of 0.506 estimated by Lunde et al (2007) using data from Norwegian birth records. Like Lunde et al, I also condition on full-term births. A large and growing literature has documented that birth weight is strongly associated with later life health and socioeconomic outcomes. Many researchers have cautioned, however, that care should be taken in interpreting this association (e.g. Gluckman and Hanson, 2005). Birth weight is a very rough proxy for fetal health and could reflect a potentially wide variety of underlying mechanisms, only some of which may be amenable to policy. Nonetheless, the results here suggest that a reasonably large fraction of health inequality among families is present at the beginning of life.

I find a somewhat smaller correlation (0.38) in the length of gestation with some strongly suggestive evidence of a difference by gender. The correlation in gestation length among sisters (0.42) is a bit higher than that for brothers (0.35). In combination, the results on birth weight and gestation suggest that an important source of the variation among families in birth weight is due to intrauterine growth retardation (IUGR) --that birth weight is lower for a given gestation length.

Interestingly I find a reasonably large correlation of about 0.3 among siblings in the parent or caregiver report of relative health at birth on a 1 to 3 scale. On the one hand, this may be viewed as quite high given that it is such a blunt measure; on the other hand, this may reflect a systematic bias in reporting by the parent or caregiver. I find lower, but still statistically significant, correlations in the incidence of poor health at birth (0.12) and in the probability of being admitted to a NICU (0.18). However both of these outcomes are statistically significant.

The columns on the right of Table 1 recalculate the overall sibling correlation (pooling both sexes) but add an additional covariate measuring family background. For example, the sibling correlation in birth weight is reduced to 0.50 from 0.53 (or by about 6 percent) when I include the race of the child. A much smaller reduction is found when I control for a five year average of family income measured in the years preceding and including birth. Similarly, including the years of completed schooling of the household head and wife (if present) does little to affect the estimate. Average health status of the household head and wife over a five years span also has little effect. Finally, the last column of the table includes all of the four sets of covariates simultaneously. This specification also does little to explain birth weight beyond what can be explained by race alone.

Looking across the other outcomes, the results of the decomposition are broadly similar. However, for gestation, the inclusion of family background factors has literally no impact whatsoever. This suggests that the observed family background characteristics may in fact play a greater role in explaining differences in IUGR than in explaining overall birth weight. It is worth noting that the literature on fetal origins of adult health and disease is focused on how environment influences affect IUGR as opposed to gestation length so this result is consistent with that literature. For health at birth, poor health at birth and NICU, the observed family background characteristics only explain between 3 and 4 percent of the sibling correlation

Childhood Health Outcomes

In Table 2 I show the results for a set of health outcomes measured during childhood. I start by estimating sibling correlations in the general health status of children. As discussed earlier, this is reported by their parents or caregivers on a 1 to 5 scale. I find that correlations vary a bit depending upon the age at which they are measured ranging from around 0.35 to 0.45. By gender, the estimates are a bit noisy suggesting some caution in interpreting the

estimates. With that caveat in mind there is suggestive evidence that the sibling correlation may rise with age for boys but not for girls.

The correlation in having a health limitation is relatively low at 0.16 but this may reflect the low incidence and relative bluntness of this outcome. The overall sibling correlation in having a learning disability is about 0.3 but the point estimate for boys is quite high at 0.50 compared to girls at 0.17. This is somewhat consistent with the pattern for development delays shown later in the table. Among the other specific health outcomes shown, correlations range from a low of 0.14 for diabetes to a high of 0.29 for allergies. One striking finding that is that the correlation in diabetes among sisters is extremely high at 0.58 with a standard error of just 0.03.

I also estimate the sibling correlation in two physical characteristics, height and weight. The overall correlation in height is 0.38 with a brother correlation of 0.49 and a sister correlation of 0.35. The brother correlation is virtually identical to the 0.492 correlation reported by Mazumder (2008) using a sample of adults in the NLSY. The sister correlation, however, is a bit lower than the 0.467 reported by Mazumder (2008). For weight, the estimated correlation is about 0.35 with a higher correlation among sisters (0.38) than brothers (0.20). Mazumder (2008) reported correlations of 0.33 for brothers and 0.29 for sisters.

The decomposition of the sibling correlation in general health outcomes during childhood by observable family characteristics appears to bear more fruit than the analogous exercise for birth outcomes. I find that nearly 36 percent of the sibling correlation in health status measured between the ages of 6 and 10 can be explained by the observed covariates. Both family income and parent health status appear to account for large portions of the sibling correlation. For health status measured both earlier and later life, however, these characteristics can account for about 15 percent of the sibling correlation. To the extent that this reflects a causal relationship, the finding that family income around the time a child enters school may matter for health may offer some hope that, for example, income support policy may play a role. For the more specific health outcomes shown in Table 3, however, it is far less clear that family income can play much of a role.

Childhood Educational Outcomes

Table 3 presents estimates for sibling correlations related to educational measures. The first measure is whether parents expect their child to attend college. As might be expected, the sibling correlation is quite large at about 0.75. Nevertheless, it suggests that some parents may distinguish between their children's likelihood of academic success. Similarly, the correlation in having less than 10 books is perhaps unsurprisingly high at 0.61. These variables may be useful as explanatory variables in models of later life outcomes in future research that tracks these individuals in their adult years.

The remaining measures in Table 3 consist of the four Woodcock-Johnson tests and the WISC digit span test. The sibling correlations in these outcomes are measured using three different samples. The first two samples limit the age range to either 6 to 10 or 11 to 15. Due to the timing of the CDS interviews this approach will by construction contain only one observation per sibling. The third approach is to allow for repeated observations on siblings and to allow the σ^2_v term to remove any transitory variation, or noise, in the data under the assumption that the object of interest is the overall performance during all ages of childhood. This approach is analogous to trying to capture "permanent income."

For the letter word score I estimate the sibling correlation to be 0.41 using just 6 to 10 year olds. This climbs to 0.49 when I examine 11 to 15 year olds. Finally, when using repeated scores of siblings the estimated sibling correlation rises to 0.62. This affirms the notion that

a large fraction of the overall variance in test scores is due to variation within individual students. Roughly similar patterns are found with the other Woodcock Johnson tests across the three specifications, though the estimates are generally smaller.¹⁰ Interestingly, in nearly every case the estimates of the correlation are larger for brothers than for sisters. Mazumder (2008) found a sibling correlation of 0.62 in military test scores using the NLSY but found nearly identical estimates by gender.

The decompositions suggest that family background and parental characteristics explain between 20 and 40 percent of the sibling correlation. For many of the outcomes family income appears to be the predominant factor. With respect to parents' expectations of a child going to college and the number of books owned, the decompositions can explain 5 percent and 9 percent of the sibling correlation, respectively, with family income being the key factor.

Young Adult Outcomes

In Table 4, I show estimates of the sibling correlations in measures from the TA modules fielded in 2005 and 2007 to older cohorts. The sibling correlation in completing high school is estimated to be 0.36, with a much higher estimate among sisters (0.50) than brothers (0.23). I obtain a roughly similar estimate for the pooled sample for attending college (0.35). These estimates, however, do not measure final educational attainment which is shown later using the adult sample. For high school grade point average the sibling correlation is estimated to be 0.32 on average and nearly three times higher for sisters (0.60) than for brothers (0.21).

The next set of estimates use composite measures produced in the TA survey. The sibling correlation in mental health is relatively high at 0.43 but is much lower for emotional well being at 0.19. The latter estimate is consistent with sibling correlations of around 0.25 for self-esteem found by Mazumder (2008). Finally, I estimate sibling correlations in the use of three drugs (two of which are illegal): diet pills, amphetamines and marijuana. The sibling correlations in all three measures are similar ranging from between 0.1 and 0.2. Interestingly, the correlation in use among sisters is consistently higher and ranges from 0.2 to 0.3. Mazumder (2008) also found a higher correlation in illegal drug use among sisters than brothers.

On the whole, the decompositions with these outcomes do not appear to be all that revealing as the estimates vary considerably across outcomes. One exception is college attendance which appears to be strongly affected by both parent education and family income.

Adult Outcomes

Table 5 presents estimates of the sibling correlations in adult outcomes using siblings drawn from the original 1968 families. For years of education I find results largely in line with previous results reported in Mazumder (2008) and Solon et al (2002) of sibling correlations in the 0.5 to 0.7 range. I also confirm Mazumder's (2008) finding of a sibling correlation in men's log earnings and log wages of around 0.5. This adds further confirmation that the previous findings of sibling correlations in earnings and wages of around 0.4 using earlier PSID samples may have been somewhat too low due to the younger age of the samples. Sibling correlations in log annual hours, however, are significantly lower and are close to zero for women. In general, there are often sharp differences in correlations in economic outcomes that depend on gender suggesting that differing patterns in labor force participation can cloud the interpretation of the results. Log family income, however, may

¹⁰The calculation test was not repeated within individuals.

offer a more comparable measure of overall socioeconomic status that takes into account the economic success of spouses. Indeed, for this measure the sibling correlations are nearly identical within the gender pairs.¹¹

With respect to health outcomes such as self-reported health and disability, I find correlations that range from 0.2 and 0.4. Combined with the estimates shown in Table 2, there is some suggestive evidence that sibling correlations in health status decline over the life course, though this may only be true for men.

Finally, using a limited set of key covariates of the parents, I find that for many of the outcomes, race, family income and parent education levels (especially the latter two) can account for portions of the correlation in outcomes among men. For example, including a five year average of log family income from the parent generation lowers the residual sibling correlation in men's wages from 0.50 to 0.39 or explaining about 22 percent of the correlation. Parent education has an especially pronounced effect on the correlation in adult health status among brothers, lowering the correlation from 0.23 to 0.14 or explaining about 40 percent. Interestingly, none of the family background measures appear to explain the sibling correlation in disability.

VI. Conclusion

By using sibling correlations, this study provides a set of new descriptive facts concerning disparities in health and socioeconomic status that emerge over the life course due to differing family and community influences. I find that sibling correlations in a range of health and socioeconomic outcomes start quite high at birth and remain high over the life course. The sibling correlations in birth weight and in adult men's wages are both estimated to be about 0.5. The sibling correlations in measures of cognitive skill during childhood are even higher.

Although the study is primarily focused on presenting a broad set of estimates of sibling correlations, statistical decompositions with a limited set of covariates representing a few key aspects of family background are also used. These results suggest if the underlying relationships have a causal component, then there may be some scope for interventions that may ameliorate the gradients in health and SES. For example, it may be possible that improvements in family income when children are between the ages of 6 and 10 may reduce disparities in cognitive and non-cognitive skills. Differences in family income also appear to account for a sizable fraction of the correlation in log wages and log family income among brothers.

It may be fruitful for future researchers to utilize this decomposition approach with more detailed data to better understand which family background characteristics matter most in explaining the inequality among families in health and SES. Further, due to data limitations, the estimates presented here do not present the cumulative effects of many factors such as early life health on long-term outcomes, a potentially important topic for future research.

References

- Altonji, JG.; Dunn, TA. Relationships among the Family Incomes and Labor Market Outcomes of Relatives. In: Ehrenberg, RG., editor. *Research in Labor Economics*. Vol. Volume 12. JAI Press; Greenwich, Conn.: 1991. p. 269-310.

¹¹Interestingly, the overall correlation among all siblings (0.39) is lower than the within gender group correlations (0.45 to 0.46) suggesting that the correlations are lower for families with siblings of both sexes.

- Altonji JG, Dunn TA. An Intergenerational Model of Wages, Hours and Earnings. *Journal of Human Resources*. 2000; 35(2):221–257.
- Ashenfelter O, Zimmerman D. Estimates of the Returns to Schooling from Sibling Data: Fathers, Sons and Brothers. *Review of Economics and Statistics*. 1997; 79(1):1–9.
- Bjorklund A, Eriksson T, Jantti M, Raaum O, Osterbacka E. Brother Correlations in Earnings in Denmark, Finland, Norway and Sweden Compared to the United States. *Journal of Population Economics*. 2002; 15(4):757–772.
- Bjorklund A, Jantti M, Lindquist M. Family background and income during the rise of the welfare state: trends in brother correlations for Swedish men born 1932–1968. *Journal of Public Economics*. 2009; 93(5–6):671–680.
- Bjorklund A, Lindahl L, Lindquist M. What More Than Parental Income, Education and Occupation? An Exploration of What Swedish Siblings Get from Their Parents. *B.E. Journal of Economic Analysis & Policy*. 2010; 10(1) (Contributions), Article 102.
- Cunha F, Heckman J. The Technology of Skill Formation. *American Economic Review*. 2007; 97(2): 31–47.
- Corcoran M, Jencks C, Olneck M. The Effects of Family Background on Earnings. *American Economic Review*. 1976; 66(2):430–435.
- Currie J. Healthy, Wealthy and Wise: Socioeconomic Status, Poor Health in Childhood, and Human Development. *Journal of Economic Literature*. 2009; 47(1):87–122.
- Gluckman, P.; Hanson, M. *The Fetal Matrix: Evolution, Development and Disease*. Cambridge, England: Cambridge University Press; 2005.
- Grossman M. Education and nonmarket outcomes. *National Bureau of Economic Review*. 2005 Aug. working paper, No. 11582.
- Levine D, Mazumder B. The Growing Importance of Family: Evidence from Brother's Earnings. *Industrial Relations*. 2007; Vol 46(1):7–21.
- Lindahl L. Does the Childhood Environment Matter for School Performance, Education and Income? - Evidence from a Stockholm Cohort. *Journal of Economic Inequality*. (forthcoming).
- Lunde A, Melve K, Gjessing H, Skjaeren R, Irgens L. Genetic and Environmental Influences on Birth Weight, Birth Length, Head Circumference, and Gestational Age by Use of Population-based Parent-Offspring Data. *American Journal of Epidemiology*. 2007; 165(7):734–741. [PubMed: 17311798]
- Mazumder B. Sibling Similarities and Economic Inequality in the US. *Journal of Population Economics*. 2008; 21:685–701.
- Page M, Solon G. Correlations between Brothers and Neighboring Boys in Their Adult Earnings: The Importance of Being Urban. *Journal of Labor Economics*. 2003; 21(4):831–856.
- Raaum O, Salvanes K, Sørensen E. The Neighborhood Is Not What It Used to Be. *Economic Journal*. 2006; 116(1):200–222.
- Schnitzlein, D. SOEP papers on Multidisciplinary Panel Data Research no. 365, Berlin. 2011. How important is the family? Evidence from sibling correlations in permanent earnings in the US, Germany and Denmark.
- Solon G, Corcoran M, Gordon R, Laren D. A Longitudinal Analysis of Sibling Correlations in Economic Status. *Journal of Human Resources*. 1991; 26(3):509–534.
- Solon G, Page M, Duncan G. Correlations Between Neighboring Children in their Subsequent Educational Attainment. *Review of Economics and Statistics*. 2000; 82(3):383–392.

Table 1

Sibling Correlation in Birth Outcomes

Outcome	All Sibs Controlling for ...							
	All Sibs	Brothers	Sisters	Race	Family Income	Parent Educ.	Parent Health	All Factors
Birth Weight	0.528 (0.019)	0.507 (0.034)	0.522 (0.031)	0.500 (0.019)	0.519 (0.019)	0.522 (0.019)	0.512 (0.019)	0.499 (0.019)
Gestation	0.377 (0.009)	0.348 (0.040)	0.424 (0.036)	0.377 (0.022)	0.376 (0.022)	0.377 (0.022)	0.377 (0.022)	0.376 (0.022)
Health at Birth	0.312 (0.023)	0.321 (0.040)	0.309 (0.039)	0.311 (0.023)	0.309 (0.023)	0.304 (0.023)	0.305 (0.023)	0.298 (0.023)
Poor Health at Birth	0.121 (0.027)	0.184 (0.046)	0.081 (0.044)	0.121 (0.027)	0.120 (0.027)	0.118 (0.027)	0.121 (0.027)	0.117 (0.027)
NICU	0.184 (0.027)	0.194 (0.052)	0.192 (0.046)	0.184 (0.027)	0.180 (0.027)	0.180 (0.027)	0.180 (0.027)	0.177 (0.027)
	3215	1640	1575	3215	3215	3215	3215	3215

Table 2

Sibling Correlation in Children's Health Outcomes

Outcome	All Sibs Controlling for							
	All Sibs	Brothers	Sisters	Race	Family Income	Parent Educ.	Parent Health	All Factors
Health Status (0 to 5)	0.438 (0.047)	0.352 (0.106)	0.534 (0.078)	0.418 (0.049)	0.397 (0.052)	0.408 (0.052)	0.380 (0.051)	0.372 (0.052)
Health Status (6 to 10)	1037	542	495	1037	1037	1037	1037	1037
	0.345 (0.029)	0.407 (0.047)	0.297 (0.059)	0.313 (0.030)	0.290 (0.031)	0.317 (0.030)	0.236 (0.033)	0.220 (0.034)
Health Status (11 to 22)	2105	1076	1029	2105	2105	2105	2105	2105
	0.368 (0.016)	0.428 (0.023)	0.371 (0.025)	0.360 (0.016)	0.335 (0.017)	0.350 (0.016)	0.319 (0.017)	0.308 (0.017)
Health Limitation	4220	2144	2076	4220	4220	3223	3223	3223
	0.162 (0.027)	0.099 (0.058)	0.077 (0.054)	0.160 (0.028)	0.154 (0.028)	0.161 (0.027)	0.147 (0.028)	0.141 (0.029)
Learning Disability (8 to 14)	3246	1656	1590	3246	3246	3246	3246	3246
	0.315 (0.042)	0.501 (0.066)	0.168 (0.087)	0.297 (0.044)	0.315 (0.042)	0.308 (0.043)	0.322 (0.042)	0.294 (0.045)
Speech Impairment	1449	728	721	1449	1449	1449	1449	1449
	0.116 (0.027)	0.170 (0.053)	0.241 (0.038)	0.112 (0.027)	0.115 (0.027)	0.117 (0.027)	0.108 (0.027)	0.101 (0.027)
Emotional Problem	3246	1656	1590	3246	3246	3246	3246	3246
	0.275 (0.024)	0.216 (0.051)	0.372 (0.038)	0.262 (0.024)	0.271 (0.024)	0.274 (0.024)	0.273 (0.024)	0.258 (0.024)
Allergy	3246	1656	1590	3246	3246	3246	3246	3246
	0.291 (0.024)	0.372 (0.040)	0.184 (0.051)	0.287 (0.024)	0.289 (0.024)	0.289 (0.024)	0.287 (0.024)	0.284 (0.024)
Anemia	3246	1656	1590	3246	3246	3246	3246	3246
	0.267 (0.023)	0.282 (0.047)	0.382 (0.039)	0.260 (0.024)	0.264 (0.023)	0.261 (0.024)	0.260 (0.024)	0.255 (0.024)
	3246	1656	1590	3246	3246	3246	3246	3246

Outcome	All Sibs Controlling for									
	All Sibs	Brothers	Sisters	Race	Family Income	Parent Educ.	Parent Health	All Factors		
Asthma	0.190 (0.026)	0.283 (0.042)	0.128 (0.047)	0.188 (0.026)	0.185 (0.026)	0.187 (0.026)	0.187 (0.026)	0.184 (0.026)		
Dev. Delays	3245 0.157 (0.027)	1656 0.255 (0.050)	1589 0.147 (0.045)	3245 0.152 (0.027)	3245 0.154 (0.027)	3245 0.158 (0.027)	3245 0.153 (0.027)	3245 0.139 (0.027)		
Diabetes	3246 0.136 (0.027)	1656 0.200 NA	1590 0.584 (0.032)	3246 0.136 (0.027)	3246 0.136 (0.027)	3246 0.135 (0.027)	3246 0.136 (0.027)	3246 0.138 (0.027)		
Hyperactivity	3245 0.172 (0.026)	1655 0.201 (0.048)	1590 0.092 (0.042)	3246 0.167 (0.027)	3246 0.166 (0.027)	3245 0.172 (0.026)	3245 0.161 (0.027)	3245 0.155 (0.027)		
Height	3245 0.375 (0.022)	1656 0.491 (0.034)	1589 0.345 (0.041)	3245 0.374 (0.022)	3245 0.375 (0.022)	3245 0.376 (0.022)	3245 0.373 (0.022)	3245 0.374 (0.022)		
Weight	2979 0.346 (0.027)	1514 0.204 (0.075)	1465 0.379 (0.053)	2979 0.339 (0.027)	2979 0.345 (0.027)	2979 0.346 (0.027)	2979 0.345 (0.027)	2979 0.341 (0.027)		
	2722	1404	1318	2722	2722	2722	2722	2722		

Table 3

Sibling Correlation in Children's Educational Outcomes

Outcome	All Sibs Controlling for							
	All Sibs	Brothers	Sisters	Race	Family Income	Parent Educ.	Parent Health	All Factors
Expected to go to College	0.749 (0.011)	0.780 (0.018)	0.733 (0.021)	0.743 (0.011)	0.712 (0.013)	0.727 (0.012)	0.732 (0.012)	0.710 (0.013)
Own fewer than 10 Books	0.608 (0.017)	0.591 (0.031)	0.653 (0.028)	0.563 (0.019)	0.554 (0.019)	0.577 (0.018)	0.568 (0.019)	0.550 (0.019)
Letter Word Score (Age 6-10)	0.411 (0.028)	0.453 (0.049)	0.325 (0.052)	0.380 (0.029)	0.326 (0.031)	0.357 (0.029)	0.366 (0.029)	0.317 (0.031)
Letter Word Score (Age 11-15)	1888 (0.023)	962 (0.039)	926 (0.041)	1888 (0.023)	1888 (0.026)	1888 (0.025)	1888 (0.026)	1888 (0.027)
Letter Word Score (All)	2367 (0.026)	1200 (0.049)	1167 (0.048)	2367 (0.027)	2367 (0.030)	2367 (0.029)	2367 (0.029)	2367 (0.031)
Passage Comp. (Age 6-10)	0.315 (0.035)	0.275 (0.074)	0.296 (0.067)	0.281 (0.037)	0.223 (0.039)	0.244 (0.038)	0.244 (0.039)	0.196 (0.040)
Passage Comp. (Age 11-15)	1731 (0.024)	882 (0.040)	849 (0.047)	1731 (0.025)	1731 (0.027)	1731 (0.027)	1731 (0.027)	1731 (0.029)
Passage Comp. (All)	0.422 (0.033)	0.494 (0.059)	0.325 (0.062)	0.405 (0.034)	0.327 (0.038)	0.337 (0.038)	0.335 (0.037)	0.286 (0.041)
Calculation (Age 6-10)	2360 (0.062)	1195 (0.108)	1165 (0.126)	2360 (0.062)	2360 (0.066)	2360 (0.069)	2360 (0.068)	2360 (0.070)
	0.617 (0.062)	0.711 (0.108)	0.472 (0.126)	0.609 (0.062)	0.500 (0.066)	0.528 (0.069)	0.524 (0.068)	0.464 (0.070)
	770	381	389	770	770	770	770	770

Outcome	All Sibs Controlling for							
	All Sibs	Brothers	Sisters	Race	Family Income	Parent Educ.	Parent Health	All Factors
Calculation (<i>Age 11-15</i>)	0.153 (0.118)	0.679 (0.158)	0.090 (0.149)	0.110 (0.121)	0.168 (0.105)	0.102 (0.118)	0.068 (0.137)	0.077 (0.127)
Applied Problems (<i>Age 6-10</i>)	520 (0.368)	265 (0.465)	255 (0.315)	520 (0.309)	520 (0.266)	520 (0.298)	520 (0.304)	520 (0.235)
Applied Problems (<i>Age 11-15</i>)	1883 (0.030)	960 (0.051)	923 (0.059)	1883 (0.032)	1883 (0.034)	1883 (0.032)	1883 (0.033)	1883 (0.035)
Applied Problems (<i>Age 11-15</i>)	0.490 (0.022)	0.566 (0.036)	0.403 (0.043)	0.462 (0.023)	0.366 (0.027)	0.391 (0.026)	0.397 (0.026)	0.330 (0.028)
Applied Problems (<i>All</i>)	2364 (0.626)	1194 (0.732)	1165 (0.552)	2364 (0.615)	2364 (0.495)	2364 (0.540)	2364 (0.534)	2364 (0.468)
Digit Span Test (<i>Age 6-10</i>)	4247 (0.027)	2159 (0.041)	2088 (0.054)	4247 (0.028)	4247 (0.032)	4247 (0.031)	4247 (0.031)	4247 (0.034)
Digit Span Test (<i>Age 6-10</i>)	0.324 (0.031)	0.352 (0.056)	0.312 (0.057)	0.302 (0.032)	0.297 (0.032)	0.291 (0.033)	0.281 (0.033)	0.258 (0.035)
Digit Span Test (<i>Age 11-15</i>)	1978 (0.410)	1003 (0.418)	975 (0.359)	1978 (0.411)	1978 (0.372)	1978 (0.384)	1978 (0.377)	1978 (0.359)
Digit Span Test (<i>All</i>)	2244 (0.026)	1116 (0.043)	1128 (0.049)	2244 (0.026)	2244 (0.027)	2244 (0.027)	2244 (0.027)	2244 (0.028)
Digit Span Test (<i>All</i>)	0.623 (0.043)	0.592 (0.060)	0.671 (0.066)	0.617 (0.034)	0.585 (0.036)	0.594 (0.036)	0.581 (0.036)	0.562 (0.037)
	4222	2119	2103	4222	4222	4222	4222	4222

Table 4

Sibling Correlation in Young Adult Outcomes

Outcome	All Sibs Controlling for						
	All Sibs	Brothers	Sisters	Family Income	Parent Educ.	Parent Health	All Factors
At least High School (<i>18 and older</i>)	0.359 (0.048)	0.234 (0.143)	0.499 (0.065)	0.297 (0.051)	0.317 (0.051)	0.327 (0.050)	0.292 (0.051)
Any College (<i>18 and older</i>)	0.351 (0.050)	0.328 (0.123)	0.351 (0.090)	0.233 (0.059)	0.227 (0.063)	0.273 (0.056)	0.192 (0.063)
High School GPA	0.323 (0.064)	0.212 (0.161)	0.596 (0.079)	0.299 (0.066)	0.298 (0.064)	0.355 (0.060)	0.325 (0.063)
Mental Health	0.432 (0.078)	0.221 (0.260)	0.631 (0.126)	0.423 (0.079)	0.393 (0.083)	0.409 (0.081)	0.388 (0.083)
Emotional Well Being	0.192 (0.109)	0.375 (0.233)	0.162 (0.246)	0.186 (0.111)	0.145 (0.118)	0.138 (0.119)	0.128 (0.122)
Ever Smoke	0.162 (0.062)	0.019 (0.189)	0.362 (0.085)	0.164 (0.061)	0.161 (0.061)	0.169 (0.061)	0.171 (0.061)
Diet Pills	0.181 (0.056)	0.107 (0.120)	0.290 (0.090)	0.182 (0.056)	0.181 (0.056)	0.181 (0.056)	0.181 (0.056)
Amphetamines	0.132 (0.059)	0.189 (0.152)	0.242 (0.084)	0.111 (0.061)	0.105 (0.062)	0.109 (0.062)	0.088 (0.064)
Marijuana	0.197 (0.060)	0.102 (0.182)	0.227 (0.108)	0.185 (0.062)	0.193 (0.061)	0.185 (0.062)	0.198 (0.061)
	1155	551	604	1155	1155	1155	1155

Table 5

Sibling Correlation in Adult Outcomes

Outcome	Brothers Controlling for						
	All Sibs	Brothers	Sisters	Race	Family Income	Parent Educ.	All Factors
Education	0.543 (0.024)	0.665 (0.031)	0.527 (0.036)	0.662 (0.031)	0.584 (0.037)	0.576 (0.038)	0.550 (0.040)
Log Earnings	1447 0.207 (0.027)	671 0.505 (0.038)	776 0.140 (0.048)	671 0.477 (0.039)	671 0.406 (0.044)	669 0.436 (0.043)	669 0.388 (0.045)
Log Wages	19531 0.391 (0.027)	10201 0.500 (0.038)	9330 0.406 (0.045)	10201 0.474 (0.039)	10201 0.385 (0.045)	10165 0.412 (0.044)	10165 0.362 (0.047)
Log Annual Hours	14861 0.040 (0.026)	7700 0.260 (0.073)	7161 0.063 (0.049)	7700 0.230 (0.077)	7700 0.241 (0.075)	7675 0.259 (0.073)	7675 0.239 (0.075)
Log Family Income	20325 0.386 (0.024)	10643 0.461 (0.039)	9682 0.452 (0.035)	10643 0.420 (0.041)	10643 0.346 (0.045)	10607 0.359 (0.045)	10607 0.302 (0.048)
Health Status	23252 0.250 (0.028)	11366 0.232 (0.049)	11886 0.353 (0.041)	11366 0.228 (0.050)	11366 0.191 (0.052)	11322 0.140 (0.334)	11322 0.143 (0.055)
Disability	13631 0.178 (0.028)	6764 0.221 (0.048)	6867 0.288 (0.046)	6764 0.222 (0.049)	6764 0.219 (0.049)	6737 0.221 (0.049)	6737 0.220 (0.049)
	14520	7077	7443	7077	7077	7050	7050