This is the accepted manuscript (post-print version) of the article. The content in the accepted manuscript version is identical to the final published version, although typography and layout may differ.

How to cite this publication
Please cite the final published version:


Publication metadata

| Title: | Intergenerational health mobility: Evidence from Danish registers. |
| Author(s): | Andersen, Carsten Berthram. |
| Journal: | Health Economics. |
| DOI/Link: | 10.1002/hec.4433 |
| Document version: | Accepted manuscript (post-print) |
| Document license: | |

**General Rights**
Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognize and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

If the document is published under a Creative Commons license, this applies instead of the general rights.
Intergenerational Health Mobility: Evidence from Danish Registers

Abstract

This paper is one of the first to investigate mobility in overall health using high-quality administrative data. The attractiveness of this approach lies in objective health measures and large sample sizes allowing twin analyses.

I operationalise health mobility by a variety of statistics: rank-rank slopes, intergenerational correlations and sibling and identical twin correlations. I find rank-rank slopes and intergenerational correlations in the range 0.11-0.15 and sibling correlations in the range 0.14-0.20. Mobility in health is thus relatively high, both when compared to similar US-based studies, and when contrasted with outcomes such as educational attainment and income.

Comparing sibling and identical twin correlations with parent-child associations confirms earlier findings in the literature on equality of opportunity, namely that sibling correlations capture far more variation than traditional intergenerational correlations. I conclude that 14-38 percent of the variation in individual health outcomes can be attributed to family background and genes, factors which the individual cannot be held accountable for. This finding suggests that simple parent-child associations may be a poor metric for measuring health mobility.

Keywords: Health, inequality, intergenerational mobility, sibling correlations, family background
1 Introduction

Inequality, and whether it is perpetuated in the family, is important both for policy makers and the general public. Most would agree that society should aim for a situation in which individuals’ prospects in life are not heavily, if at all, constrained by their family background. In other words, policy makers should pursue equality of opportunity.

In this paper I estimate to what extent health is passed down through generations and reproduced within the family in a Danish context. Exploiting rich register data I construct an intergenerational sample where adult health can be measured in two consecutive generations. I am able to identify not only parent-child links but also siblings and twins. This allows me to quantify how much traditional mobility studies miss by only studying the transmission between parents and offspring relative to a broader emphasis on the general importance of family background which sibling and twin correlations represent.

Economists have long been interested in the transmission of status across generations. Research on intergenerational mobility in traditional outcomes such as income or earnings (Björklund and Jäntti, 2012; Chetty et al., 2014; Jantti et al., 2006; Landersø and Heckman, 2017; Solon, 1992; Zimmerman, 1992), and wealth (Boserup et al. (2016)) has received a lot of attention whereas sociologists have been more concerned with transmission of class and occupation status (Erikson and Goldthorpe, 1992; Torche, 2015).

However, the transmission of health status remains severely understudied. This is unfortunate for several reasons. In itself, health is arguably central to human well-being: For example, Grossman (1972) develops a model in which health enters the utility function directly as something individuals derive pleasure from. Moreover, health impacts how individuals fare on the labour market. Health shocks induce retirement (Datta Gupta et al., 2015; Dwyer and Mitchell, 1999; McClellan, 1998) and lead to unemployment and lower income (Garci´a-Gómez et al. (2013)). In the Grossman model worsening health decreases income as there is less productive time available (Grossman (1972)). Thus, knowing to what extent health outcomes are reproduced in families should be important for policy makers and for normative discussions of fairness (Arneson (2018)).

Expanding the focus of intergenerational mobility from traditional labour market
outcomes to health outcomes is challenging. It is less obvious how to measure health compared to income. The literature analysing parent-child health links has seen different approaches, where examples include body mass index (Dolton and Xiao (2017)), birth weight (Currie and Moretti (2007)), longevity (Parman et al. (2010)), mortality-weighted hospitalisation visits (Björkegren et al. (2019)), principal components from self-reported ailments (Halliday and Mazumder (2017)) and time-averages of self-reported health (Graeber, 2018; Halliday et al., 2021; Mazumder, 2011).

My contribution to this literature lies in (i) a health measure based on rich administrative data (ii) going beyond simple parent-child associations and quantifying equality of opportunity by sibling and twin correlations. I measure health by taking the first principal component of a battery of objective health utilisation measures relating to hospital and doctor visits. I argue that it is reasonable to rely on health care utilisation in a context where access is free and universal. I show that the health measure meaningfully correlates with observables such as mortality and transition to disability pension. My health measure is also relatively robust to attenuation bias and lifecycle bias often encountered in studies on intergenerational mobility.

One significant advantage of developing a measure based on register data compared to, say, self-reported health, is the large sample sizes then being available for analysis. Moreover, the register data covers an entire population of Danes. Finally, although measures as self-reported health have been found to predict mortality quite well, the nature of such health assessments represent an inherent subjectivity which administrative data on health care utilisation obviously are not prone to.

Methodologically, I supplement traditional parent-child associations with sibling correlations and identical twin correlations, for the latter using a method developed in Nicoletti and Rabe (2013). The rationale for this exercise is promoted in Björklund and Jäntti (2009, 2012) but rarely used in the literature, namely that sibling correlations and identical twin correlations constitute plausible lower and upper bounds, respectively, on the total importance of family background.¹

¹This is possibly due to the heavy data requirements. Absent a twin database this methodology requires large sample sizes as twins only comprise 1-2% of the general population.
This study, as most within the intergenerational mobility literature, is deliberately descriptive in nature as the aim simply is to describe the extent of transmission and reproduction of general health within families. What do we learn from this? First, by comparing estimates of mobility across countries and settings we may get indirect evidence on the importance of institutions and policies, where examples include income inequality, income mobility, access to education and universal health care insurance. For mobility in income and earnings this aspect has been stressed in discussion of "The Great Gatsby Curve" (Corak (2013)) as well as the large variation in mobility levels within US commuting zones (Chetty et al. (2014)). Also, we learn something about whether to expect large causal health effects on the offspring generation from interventions aimed at the parent generation.

I find rank-rank slopes in the range 0.11-0.15 which is approximately half of what is found for US data (Halliday et al., 2020, 2021), two thirds of estimates for Germany (Graeber (2018)), and on par with recent estimates from Sweden (Björkegren et al. (2019)). In contrast to most studies I supplement parent-child associations (such as the rank-rank slope) with sibling and twin correlations. I find that the parent-child approach drastically underestimates the broad importance of family background as only 9-11 percent of the health resemblance between sibling can be traced back to common parental health. I find sibling correlations in health in the range 0.14-0.20 and identical twin correlations between 0.34-0.38. Family background seems to play a larger role for women than men whereas the opposite is usually found for income mobility (e.g., Björklund and Jäntti (2012)).

Central implications from this study are that the parent-child link in health is not particularly salient relative to factors broadly shared by siblings, that health mobility is higher than income mobility, and tentatively that health mobility levels seem to be higher in countries with universal health care insurance and higher levels of economic mobility.

The rest of the paper is organized as follows. Section 2 provides some background on related studies on intergenerational mobility. Section 3 describes the data and measurement of health. Section 4 lays out the methodology to evaluate how health is reproduced in families. Section 5 presents results. Finally, Section 6 concludes and discusses implications.
Various sensitivity checks can be found in the Appendix.

2 Related Literature

The literature on the relation between health outcomes and family background goes back at least to the turn of the nineteenth century with Galton’s study of heredity of human height (Galton, 1886). However, in this review I focus mostly on studies that are characterised by one or more of the following: Addresses measurement issues as known from the income mobility literature, employs sibling correlations, or employs a broader health measure than a purely anthropometric one such as height.

The studies most comparable to mine are Halliday et al. (2021) and Graeber (2018), both explicit mobility studies that address measurement issues as lifecycle bias and attenuation bias. Halliday et al. (2021) proxy health by long-run averages of self-reported health in the PSID and find rank-rank slopes in the range 0.21-0.29 where the strongest link is found between mother and daughter. Graeber (2018) extracts a continuous one-dimensional health measure based on self-reported health measures from the German Socio-Economic Panel and obtains rank-rank slopes in the range 0.20-0.22. Halliday et al. (2020) builds on Halliday et al. (2021) and models self-reported health status in a latent variable framework which yields largely similar results for the rank-rank slopes.

Mazumder (2011) estimates sibling correlations on a wide range of outcomes in the PSID, including income, education and health, and for different ages ranges, including childhood, adolescence and in adulthood. Adult sibling correlations in self-reported health on a Likert scale are 0.23 for brothers and 0.35 for sisters. The corresponding figures for adult disability are slightly lower. Björkegren et al. (2019) focus on quantifying the relative importance of pre-birth and post-birth factors in health, and the resemblance to my study is the Scandinavian setting, the health measure, and overlap of reported statistics such as the rank-rank slope. Using Swedish register data they utilise two health measures, one based on hospitalisations (somewhat similar to the measure used in this paper), and one health measure where hospitalisations are weighted based on the five-year mortality rate associated with the specific diagnoses. For the two health indices they find rank-rank
slopes in the range 0.13-0.15.

Other (US) studies that explore intergenerational health mobility, but measure offspring health in childhood are Halliday and Mazumder (2017) and Fletcher and Jajtner (2019). Halliday and Mazumder (2017) find high sibling correlations in the range 0.5-0.6 and Fletcher and Jajtner (2019) obtain national rank-rank slopes around 0.17 with considerable geographic variation. Due to health measurements taking place at very different points in the lifecycle (and e.g., Fletcher and Jajtner (2019) only observe one year of parental health) it is not obvious whether results are readily comparable to my study.

Further, it should be taken into account that Halliday and Mazumder (2017) estimate a full latent variable model which is beyond the scope of this study. In general, estimating full latent variable models tend to deliver lower estimates of mobility than using blunt(er) proxy variables, at least for sibling correlations (e.g., compare Halliday and Mazumder (2017) to Mazumder (2011)). However, the difference seems modest for parent-child mobility measures (e.g., compare Halliday et al. (2020) to Halliday et al. (2021)).

Interestingly, both Halliday et al. (2021) and Fletcher and Jajtner (2019) tentatively find higher mobility levels for those with parents having health insurance.² Note also that only two of the above mentioned studies report sibling correlations, thus potentially only analysing the ”tip of the iceberg” with the sole focus on parent-child associations (Björklund and Jäntti (2012)).

Büglmayer and Schnitzlein (2018) find sibling correlations in physical health of 0.23 using the German Socio-Economic Panel (SOEP) and Roos et al. (2014) report comparable magnitudes of 0.26-0.32 when proxying health by hospital and physician costs in a Canadian context. Both these studies, however, are based on individuals in their youth or adolescence.

Many other studies analyse transmission of anthropometric outcomes such as height (as Galton (1886)), weight and body mass index. For Sweden, Björklund and Jäntti (2012) find a brother correlation of adult height of 0.53 which is virtually identical to what Mazumder (2008) reports for the US. For body mass index and weight, Mazumder (2008)

²It should be emphasized that Halliday et al. (2021) caveats this finding and leaves a formal exploration for future research.
estimates sibling correlations around 0.27-0.33, not too different from 0.35 for body mass index found for German adolescents (Bügemayer and Schnitzlein (2018)). Also, when gauging mobility in body mass index by parent-child associations, Dolton and Xiao (2017) find a striking resemblance across six very different countries\(^3\), all of them displaying intergenerational elasticities of around 0.20. The advantage of anthropometric measures is that they are readily available in many data sets and probably prone to negligible measurement error. However, it is less clear whether they reflect “health” variables or, say, risk factors. Also, while the outcomes are continuous it complicates the interpretation that they are not monotone, i.e., it may be considered unhealthy to be obese as well as being underweight.

Finally, some papers study how longevity depends on family background (Parman et al. (2010) and Piraino et al. (2014)). While longevity is obviously appealing as an intuitive metric of health, it involves the downside of providing very retrospective insights if one wants to analyse two consecutive generations with fully resolved life spans.

Compared to other literature, the relative merits of my study is the nationally representative sample of individuals with high-quality panel data from adulthood based on sample sizes that allow for both sibling and twin analyses.

3 Data

Below, I outline the data sources, sample selection and the chosen health measure. Measuring health, especially in an intergenerational context, is not obvious. Here, the health measure should be continuous, based on register data, and plausibly correlate with “objective” health events (e.g., mortality).

The literature on intergenerational mobility has emphasized two inherent biases in estimating the degree of persistence in a society. Lifecycle bias captures biases arising from measuring outcomes at sub-optimal ages, i.e., in certain points in their lifecycle that does not accurately proxy their lifetime outcome (Grawe (2006)). For example, measuring

\(^3\)Countries are sampled to be at very different stages of economic development: China, Indonesia, England, Spain, Mexico and the US.
lifetime earnings by a multi-year average of earnings while individuals are in their 20s is probably misleading. *Attenuation bias* reflects a bias arising from measurement error in only including, say, single-year measurement earnings to capture lifetime income or health. Intuitively, there will be substantial noise in such estimates, leading to an attenuation of the estimated parameters.

The health measure I extract fares reasonably well on these points. Due to space limitations, not all details can be provided here. Below, the reader is in some instances referred to the Appendix or *Andersen et al. (2019)* for more in-depth explorations of the health measure.

### 3.1 Data Sources and Sample Selection

I use Danish register data provided by Statistics Denmark spanning the years 1977-2014. All Danish registers contain a unique personal identifier allowing me to link individuals to parents and across different registers. The data has information on general practitioner visit (GP visits) and in- and outpatient hospital care. Denmark has a universal health care system where most services carry little or no out-of-pocket costs for the patient.

For constructing a health measure the aim is to obtain health information on two consecutive generations, and by implication, information on sibling and twin links. On the one hand, this speaks for including as early cohort as possible as health information is revealed progressively over the life cycle. Also, to overcome measurement error (attenuation bias) it is likely misleading to include only a few years of health measurement. On the other hand it is undesirable to include very early cohorts as they to a large extent will be deceased before they enter the health registers. Further, the experience of individuals of very early cohorts may not be representative of the current day environment.

In the main specification I select 71,460 children born between 1960-1963 with parents born 1930-1940. The main sample consists of 36,178 sons from 32,674 families and 35,282 daughters from 31,983 families. The sample contains 58,535 (58,609) unique fathers.

---

4 In the income literature it is suggested to use at least 9 years of measurement to overcome this bias (Mazumder (2005)). For health, many studies fail to consider these issues. One exception is *Halliday et al. (2021)* who find that estimates tend to flatten after 8 years of included health measurements.
(mothers) and 444 male twins, 446 female twins and 430 mixed sex twins. This sample is representative of the Danish population and constitutes more than 95 percent of the population in the child cohorts. In the primary specification I measure health over long periods for both generations (14 years for children, 10 years for parents). This is a longer time span than most papers within the intergenerational (health) mobility literature.

The cohort selection and utilisation of data sources is illustrated in Figure 1, which is scaled to show the years 1918-2018, where the text above the line denotes the cohort selection and the text below the line denotes the data sources used along with the interval in calendar time.

I measure every generation at the same point in their lifecycle. For health, I measure parental health in the age interval 60-70 and child health in age 36-50. It may seem suboptimal to measure outcomes at different points in the life cycle. However, as shown in Andersen et al. (2019), for lifecycle bias and attenuation bias, results turn out to be insensitive to both particular cohort and age restrictions. For lifecycle bias, results are similar when measuring health at different points in the lifecycle while keeping the length of the measurement window fixed. For attenuation bias, rank-rank slopes stabilize when including around 10 years of health measurements. Hence, provided the measurement interval is sufficiently long, the exact placement in the lifecycle seems of less importance in this setting. I also experiment with a sample where I do not impose survival restrictions on both parents simultaneously which delivers similar estimates (Andersen et al. (2019)). Finally, when the measurement windows in the lifecycle are moved closer to each other for adjacent generations, e.g., changed to 49-54 and 60-65 for the child and parent generation respectively, results are similar (See Appendix C).

The primary estimation sample is then the subset of children in birth cohorts 1960-1963 whose parents are neither dead nor emigrated by age 70. Individuals of the child

---

5It is also a question whether cross-country comparability is invalidated by differential survival patterns across countries. For example, Halliday et al. (2021) also restrict parental survival till 70 in some specifications. However, at least for Denmark compared to the US, survival patterns appear to be on par. For example, in 2010 the Danish (US) survival rate for males at age 70 was 76 (73) percent, and the 2010 Danish (US) survival rate for females at age 70 was 85 (83) percent (Statistics Denmark, 2019; United States Social Security, 2019).
3.2 Health Measure

The aim is to extract a measure of health that is approximately continuous. A key rationale here is to make results (more) comparable to other studies in intergenerational mobility, and enabling use of e.g., the rank-rank correlation as popularized in Chetty et al. (2014).

I construct a health measure by taking the first principal component of a battery of indicator variables from health utilisation measures of hospital and general practitioner visits (the latter also referred to as doctor visits or GP visits). Principal components is a simple way to reduce high-quality health information into a one-dimensional continuous measure while retaining as much variability as possible.

On the face of it, using health care utilisation in a measure of health may seem as a conflation of two distinct concepts, namely health and health care. Indeed, Grossman (1972) makes a distinction between the two in his model where individuals demand health care as a means to producing health, which is treated as a durable capital good. Also, individuals may seek different levels of care given the same levels of underlying "true" health status.
This approach would be questionable in a setting where it is common to experience financial barriers to vital health care. For example, Figure 1 in Devaux and De Looper (2012) shows how number of doctor consultations in the US correlates positively with income, where the opposite is the case for Denmark. Doorslaer et al. (2004) shows that when controlling for health needs, there is no income-related inequity in Denmark for GP visits and the probability of seeing a medical specialist, but income-related inequity remains in the number of specialist visits.

This finding may indicate broad problems in the sense that different segments of the population utilise health care differently, conditional on health needs. However, on the other hand, this particular finding is probably less of a concern here as the hospitalisation indicators enter as binary variables, not as count variables. Still, even in Denmark health care utilisation is not a one-to-one reflection of actual underlying health. For example, Fadlon and Nielsen (2019) find that individuals increase health care consumption (e.g., GP visits) immediately after the occurrence of a health shock to someone in the family or a close coworker.

Hence, it clear that the health measure suggested here has its downsides. This is not unlike other health measures such as mortality where individuals may die for “random” reasons, and which potentially miss health inequality present while alive. Another caveat in the current setting is that the GP acts as a gatekeeper so hospital care, i.e., the diagnoses used in the health measure, is conditional on having visited the doctor. The limitations of the health measure should be kept in mind going forward.

The intuition of the health measure is simple: an individual is unhealthy the more she visits the doctor, and the more different ailments she is treated for in hospital. Another approach is to explicitly anchor the health measure to mortality as one of the health indices in e.g., Björkegren et al. (2019). However, given that inequality in healthy life years is generally larger than inequality in longevity (Brønnum-Hansen et al. (2017) and Wood et al. (2006)) it seems plausible that hospitalisation and doctor visits data capture a broader aspect of health than mortality.
3.3 Descriptive Statistics

Table 1 shows the variables going into the health measure. The first set of variables are the hospital variables. The categorization of these diagnoses are made on the basis of a 23-category aggregation of International Classification of Diseases (ICD) used by Statistics Denmark (Statistics Denmark (2011)).

This data contains individual level inpatient and outpatient hospital contacts and the duration of the stay. From this information, I construct the set of binary variables in Table 1 which indicates the fraction who have ever been registered with the given ailment, where both inpatient and outpatient care is included. Again, the intuition is that having had hospital visits within many diagnosis categories is associated with worse health. From the included diagnoses in Table 1 it may seem that they do not all capture important health events. For example, contacts within the cardiovascular category are more intuitively related to a concept of bad health than, say, diagnoses related to eye-sight or infections. Results are not sensitive to exclusion of these conditions. Ultimately, I included all of the diagnosis categories apart from sterilisations, poisonings and concussions. I also excluded the explicitly gender related categories such as pregnancy complications and gynecological conditions. Such conditions may reflect health levels as well as fertility decisions. Again, results for women are insensitive to inclusion of these.

In Table 1 it is seen that for most categories women are more often diagnosed relative to men. Also, unsurprisingly, the parent generation suffers more from e.g., cardiovascular diseases. 29 percent of fathers have been in contact with a hospital on this account some time during their sixties. Moving to the lower part of Table 1 shows the mean number of doctor visits by gender and type-specific quintile. Large differences are seen here. For example, fathers in the lowest quintile on average only have 20 doctor contacts over the course of 10 years, while this number is 255 for the highest quintile. In general, the medical literature finds that women consume more health services than males (Bertakis et al., 2000; Redondo-Sendino et al., 2006).

It does not make sense to comprehensively compare the evolution of ailments and general health care usage across generations. Secular trends in hospital diagnoses, health
and labour markets complicate interpretation of Table 1 as revealing something about differences in absolute health. This should pose no problem as the empirical specifications standardize intergenerational associations by either the intergenerational correlation or the rank-rank slope. In addition, all outcome variables are net of cohort effects. It also presents yet another argument for relying on sibling correlations over conventional approaches regressing parent status on child status. Siblings are often closely spaced in time compared to parents and offspring. Hence, secular trends broadly defined will largely be shared by siblings.

All the variables of Table 1 make up the ingredients in condensing multidimensional health register information into a one-dimensional health measure via principal components.\(^6\) For rank-rank specifications the individuals are then assigned a percentile rank within each cohort and gender according to their predicted value of the first principal component. The ambition is not to construct an elaborate factor model of a latent phenomenon, but it is rather an exercise in pragmatic dimensionality reduction. Appendix B shows scree plots and factor loadings for the health measure. In particular it is seen that the first component by far contains the most information, and subsequent components have little explanatory power.

Further, in Appendix A, I show how the extracted health measure for both generations meaningfully correlates with relatively objective health events such as transition to mortality and disability pension (Figure A.1 and Figure A.2). In particular, there is a monotone negative relationship between predicted health rank and four-year mortality (disability) rate for parents (children). This correlation is not solely driven by hospital visits as similar patterns can be seen when only including GP visits (Figure A.3-A.5).

\(^6\)‘pca’ in Stata 14.
Table 1: Descriptive Statistics of Health Care Utilisation

<table>
<thead>
<tr>
<th>Share with ailment</th>
<th>Sons</th>
<th>Daughters</th>
<th>Fathers</th>
<th>Mothers</th>
</tr>
</thead>
<tbody>
<tr>
<td>Infectious Diseases</td>
<td>.03</td>
<td>.03</td>
<td>.03</td>
<td>.03</td>
</tr>
<tr>
<td>Lung Diseases</td>
<td>.09</td>
<td>.08</td>
<td>.15</td>
<td>.14</td>
</tr>
<tr>
<td>Nervous System</td>
<td>.1</td>
<td>.11</td>
<td>.14</td>
<td>.1</td>
</tr>
<tr>
<td>Cardiovascular Diseases</td>
<td>.09</td>
<td>.09</td>
<td>.29</td>
<td>.23</td>
</tr>
<tr>
<td>Arteries and Veins</td>
<td>.02</td>
<td>.03</td>
<td>.07</td>
<td>.05</td>
</tr>
<tr>
<td>Varicose Veins</td>
<td>.01</td>
<td>.04</td>
<td>.02</td>
<td>.04</td>
</tr>
<tr>
<td>Blood Disorders</td>
<td>.01</td>
<td>.04</td>
<td>.03</td>
<td>.03</td>
</tr>
<tr>
<td>Gastrointestinal Diseases</td>
<td>.22</td>
<td>.26</td>
<td>.29</td>
<td>.27</td>
</tr>
<tr>
<td>Rheumatic Diseases</td>
<td>.68</td>
<td>.61</td>
<td>.47</td>
<td>.55</td>
</tr>
<tr>
<td>Skin Diseases</td>
<td>.11</td>
<td>.13</td>
<td>.08</td>
<td>.08</td>
</tr>
<tr>
<td>Eye Diseases</td>
<td>.16</td>
<td>.1</td>
<td>.13</td>
<td>.15</td>
</tr>
<tr>
<td>Ear, Nose and Throat</td>
<td>.12</td>
<td>.12</td>
<td>.22</td>
<td>.15</td>
</tr>
<tr>
<td>Endocrine Diseases</td>
<td>.07</td>
<td>.12</td>
<td>.14</td>
<td>.16</td>
</tr>
<tr>
<td>Mental Illnesses</td>
<td>.05</td>
<td>.05</td>
<td>.03</td>
<td>.03</td>
</tr>
</tbody>
</table>

Mean GP Visits by Quintile

<table>
<thead>
<tr>
<th>GP Visits Q1</th>
<th>GP Visits Q2</th>
<th>GP Visits Q3</th>
<th>GP Visits Q4</th>
<th>GP Visits Q5</th>
</tr>
</thead>
<tbody>
<tr>
<td>16</td>
<td>44</td>
<td>20</td>
<td>32</td>
<td></td>
</tr>
<tr>
<td>43</td>
<td>95</td>
<td>53</td>
<td>78</td>
<td></td>
</tr>
<tr>
<td>70</td>
<td>138</td>
<td>86</td>
<td>123</td>
<td></td>
</tr>
<tr>
<td>108</td>
<td>197</td>
<td>132</td>
<td>181</td>
<td></td>
</tr>
<tr>
<td>215</td>
<td>349</td>
<td>255</td>
<td>331</td>
<td></td>
</tr>
</tbody>
</table>

Observations 36,178  35,282  58,535  58,609

Notes: The upper part shows the share of individuals who have ever been in inpatient or outpatient hospital care with the given ailment, over the years of observation (36-50 for the child generation and 60-70 for the parent generation). The lower part shows the mean general practitioner (GP) visits by quintile (i.e., doctor visits). The share of individuals with zero GP visits is less than 1% for sons, daughters, father and mothers.
4 Methodology

4.1 Parent-Child Associations: The Workhorse model of intergenerational mobility

The point of departure of most of the literature on intergenerational mobility is a simple linear regression model relating outcomes of consecutive generations as

\[ Y_C = \alpha + \beta Y_P + \epsilon, \]

(1)

where \( Y_C \) and \( Y_P \) denote status of the child and the parent, respectively. Here, outcome variables are residualised on cohort effects before estimation and the error term, \( \epsilon \), represents all unobserved factors affecting child status. It is beyond the scope of this paper to investigate mechanisms, hence there is no vector of variables with e.g., education or cognitive ability. The parameter of interest is \( \beta \) which shows the strength of the association between parent and child outcomes. If positive and numerically large, mobility is low. If close to zero, intergenerational mobility is high since child status is not very dependent on parent status. In the current setup it is if course unwarranted to attach any causal interpretation to \( \beta \).

A frequent implementation of the above is the case where both \( Y \)-variables are log of income or earnings in which case \( \beta \) is referred to as the intergenerational elasticity (e.g., Chetty et al. (2014)). In the health mobility literature, \( \beta \) has been reported as the intergenerational health association (Halliday et al. (2021)).

In the results for parent-child associations I report two primary statistics, the intergenerational correlation (IGC, or \( \gamma \)) and the rank-rank slope (henceforth, \( \beta_R \)). The IGC is a natural companion statistic to the intergenerational elasticity, or intergenerational health association, and simple algebra reveals that

\[ \gamma = \frac{\sigma_P}{\sigma_C} \beta, \]

(2)

where \( \sigma_C \) and \( \sigma_P \) are the associated standard deviations of the child and parent distributions, respectively. The two mobility statistics are identical only when \( \sigma_C = \sigma_P \). In the
modal outcome variable in the literature, income, this will often lead to elasticities being higher than correlations because the outcomes of the child generation are more dispersed (that is, $\sigma_C > \sigma_P$). Thus, $\gamma$ is a statistic that accounts for the fact that inequality may differ between generations.

The second statistic, the rank-rank slope, has gained prominence in the literature recently (e.g., Chetty et al., 2014; Halliday et al., 2021). The rank-rank slope is the Spearman correlation coefficient between $Y_C$ and $Y_P$, or equivalently the Pearson correlation coefficient between the ranked measures (as well as the regression coefficient due to identical standard deviations of ranks). This measure only uses information on the relative position between individuals; hence, it makes no assumptions about the distribution of the data apart from monotonicity. It has also been found to be more robust to measurement issues such as life-cycle bias.

Another appealing feature of the rank-rank slope is that it readily compares across domains and groups (e.g., comparing rank-rank slopes in income with rank-rank slopes in health, and comparing rank-rank slopes for different ethnic groups). However, empirically it turns out that rank-rank slopes and intergenerational correlations are virtually identical for health (see the Results section).

Finally, to ease comparison with earlier studies I report the expected rank of children with parents at any percentile $p$, here $p = 25$ and $p = 75$. Assuming linearity of the rank-rank slope this is the mean health rank of children with parents in the bottom and the top half of the parental health distribution. This information of “directional” mobility has been used in studies of mobility, both income (Chetty et al. (2014)) and health (Halliday et al. (2021)).

---

7In general, the predicted health measure is not a cardinal variable given that it does not have an equidistant scale. In this sense, intergenerational mobility studies using health indices or self-reported health are different than intergenerational income studies.
4.2 Sibling and Twin Correlations: Bounds on the Total Importance of Family Background

4.2.1 Sibling Correlations

The literature on intergenerational mobility may overlook important components of family background by focusing on a single parental resource (Björklund and Jäntti, 2009; Bredtmann and Smith, 2018; Mazumder, 2008; Solon, 1999), i.e., $Y_p$. Ultimately, what should be relevant for gauging equality of opportunity is whether health status is reproduced according to general family background, not the correlation of child health with one specific parental variable (i.e., parental health). In addition to the parental resource, a sibling correlation takes into account all unobserved factors that are shared by siblings and that are uncorrelated with the parental resource. This can be written informally as:

\[
\text{Sibling correlation} = \gamma^2 + \text{other shared factors that are uncorrelated with the parental variable,}
\]

where $\gamma$ is the intergenerational correlation defined earlier. See Solon (1999) for a derivation, and Piraino et al. (2014) and Björklund and Jäntti (2012) for applications. The above equation shows how much of the importance of family background that potentially is missed by only estimating models of the kind in equation (1). The conceptual framework of the sibling correlation is simple. The outcome of the child, $Y$ (dropping the subscript from equation (1) for simplicity), can be decomposed as

\[
y_{ij} = a_i + b_{ij},
\]

where $a_i$ is shared by siblings in a family and $b_{ij}$ is the individual-level deviation from $a_i$ for individual $j$ in family $i$ (orthogonal by construction). Then, the variance of the outcome is given as

\[
\sigma_Y^2 = \sigma_a^2 + \sigma_b^2
\]

Thus, the fraction of the variance in the outcome, which is attributable to family background, is given as
\[ \rho = \frac{\sigma_a^2}{\sigma_a^2 + \sigma_b^2} \]

The sibling correlation, \( \rho \), is the correlation being two randomly picked siblings. The numerator holds the between-family variation and the denominator holds the sum of this component and the within-family variation. Intuitively, if the numerator is large relative to the denominator, that is, if the variation between families is large relative to the individual variation, then shared factors in the family is important. In other words, siblings should then resemble each other to a high degree. If \( \rho \) is zero, family background has no bearing on health outcomes of children, while a positive \( \rho \) indicates that siblings to some degree share health outcomes, thus pointing to the importance of family background.

I estimate three types of sibling correlations: Correlations between non-twin siblings, correlations between twins, and correlations between monozygotic twins (See description below and Table 2).

In all the estimations, the variance components of equation (5) are estimated by Restricted Maximum Likelihood (REML), following other papers in the literature (Björklund and Jäntti, 2012; Bredtmann and Smith, 2018; Mazumder, 2008). Similarly in line with most literature, I define siblings as individuals who share a biological mother and a biological father.\(^8\) The outcome variables for all sibling correlations are the predicted one-dimensional health measures, which is most comparable to other literature on sibling correlations. All standard errors are obtained by bootstrapping with 100 replications.

While the sibling correlation arguably is a broader measure of intergenerational mobility than regressions of the kind in equation (1), it overlooks some aspects. Examples include the fifty percent genes the siblings do not share, differential treatment by parents, and (other) time-varying factors in the upbringing. However, the advantage of the sibling correlation is that it reflects everything that is time constant within the family, observed as well as unobserved factors. Therefore, variables difficult to measure such as family values, attitudes and preferences, to the extent that they do not vary over time, are also captured

\(^8\)This is based on the registers which is of high quality with respect to parent-child linkage. However, in very few cases, the definition will instead only capture the social father.
by this broad “omnibus” measure. Hence, whereas parent-child associations capture how far the apple falls from the tree, sibling correlations measure how far the apples fall from each other. For example, Solon (1999) notes that only forty percent of US sibling correlations in long-run earnings can be explained by parental earnings. Björklund and Jäntti (2012) find that the importance of background is even more severely underestimated in a Swedish setting for a wide range of measures, including earnings, height and IQ, when estimating equations of the form in equation (1) compared to estimating sibling correlations.9

In contrast, sibling correlations offer a way of quantifying the importance of family background by exclusively focusing on individuals growing up at approximately the same point in time. Note that the offspring generation in the present analysis consists of individuals born 1960-1963. For example, sibling correlations capture the fact that the high-SES (low-SES) siblings of this generation, statistically speaking, grew up in an environment where smoking eventually became relatively less (more) predominant. The results displayed below show that in the case of health, traditional parent-child associations miss even more of the sibling resemblance than outcomes as education and income, height and IQ (Björklund and Jäntti (2012)).

9For the case of health, there may be additional historic reasons for why parent-child associations are inadequate. The case of smoking presents an interesting example of a “cross-over” of prevalence among different groups in society. It was a symbol of wealth and status in the first half of the 20th century, eventually also taken up by the middle and lower classes. After growing public awareness of its harmful effects in the 1960s the SES patterns of this practise reversed, with the low-SES groups now consuming more tobacco (Pampel, 2007; US Department of Health and Human Services, 2014). Hence, among young adults in the 1950s and early 1960s, high-SES individuals would probably smoke more than their low-SES counterparts, but their offspring would smoke less than the children of the low SES-group. The high-SES group of parents would in all likelihood still be in overall better health than the (less tobacco consuming) low-SES group, but this cross-over of preferences would tend to attenuate the correlation between parent and offspring health.
4.2.2 Identical Twin Correlations

The sibling correlation can be interpreted as a lower bound on the total importance of family background. Identical twin correlations can be seen as a corresponding upper bound (Björklund and Salvanes (2011)). In addition to what non-twin siblings share, identical twins fully share genetic endowment and shocks to the family.\footnote{Regarding genetics, even identical twins have different epigenetic profiles at age five (Mill et al. (2006). However, the bounding argument made here will for practical purposes disregard such epigenetic drift.} Plausibly, they also interact more extensively than the general population, potentially further increasing the correlation between them. This can be seen as "excessive" interaction that is misleading to generalize to the general population but it may also lead to greater trust in the identical twin correlation being a genuine upper bound.

The upper bound argument relies on assumptions that are unlikely to be completely accurate in practice. For example, even monozygotic twins may differ in birth weight and birth length (see for example Table 1 of Yokoyama et al. (2016) where an average significant difference of 30 grams is found). This may pose a problem given that e.g., birth weight predicts long-run health and labour market outcomes (Bharadwaj et al. (2018)). Also, parents may deliberately treat twins differently to set them apart (Tourrette et al. (1989)). However, it is unclear whether such parenting practice should be so pronounced as to lead to significantly different health outcomes in the long run. Ultimately, some of these assumptions are untestable. However, I find that results for non-twin siblings are unchanged whether controlling for birth order or not. Hence, in this situation of even larger average birth weight differences, neither sibling correlations nor parent-child associations are affected.

For estimation of the identical twin correlations I employ a technique from Nicoletti and Rabe (2013) and Björklund and Jäntti (2012). This technique is a way to identify identical twin correlations in situations where the data do not have information on zygosity as is the case here, but where twins can be identified. This method is described in detail in the above references, so an explicit presentation is not given here (Specifically, the reader is referred to Appendix A of Björklund and Jäntti (2012).) The key insight of the method is that twins of different sexes necessarily must be dizygotic. If one then assumes that, for
Table 2: Various forms of sibling correlations

<table>
<thead>
<tr>
<th>Name</th>
<th>Symbol</th>
<th>Description</th>
<th>Estimation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sibling correlations</td>
<td>$\rho_S$</td>
<td>Lower bound on the importance of family background.</td>
<td>Sibling identifiers available in data.</td>
</tr>
<tr>
<td>Twin Correlations</td>
<td>$\rho_T$</td>
<td>&quot;Intermediate bound&quot; on the importance of family background.</td>
<td>Twin identifiers available in data.</td>
</tr>
<tr>
<td>Monozygotic twin correlations</td>
<td>$\rho_{MZ}$</td>
<td>Upper bound on the importance of family background.</td>
<td>Zygosity information not available in data. Impose regularity conditions on the variance components (Follows procedure in Nicoletti and Rabe, 2013 and Björklund and Jäntti, 2012).</td>
</tr>
</tbody>
</table>

Notes: Table 2 shows the three types of sibling correlations estimated in this paper. Estimation is based on equation (6).

Both genders, the difference between same-sex dizygotic twins and mixed-sex dizygotic twins is identical to the difference between same sex non-twin siblings and mixed-sex non-twin siblings the correlation between monozygotic twins is identified. Thus, the identifying assumption allows different levels of variances across different genders and sibling types but restricts the differences between same-sex and different sexes to be the same in the group of non-twin siblings and the group of DZ-siblings.

Table 2 sums up the various forms of sibling correlations used in this study.

5 Results

All results are net of cohort fixed effects and presented by gender. I have experimented with specifications also controlling for birth order and the age difference between offspring and parents but found no differences.
5.1 Parent-Child Associations

Table 3 shows parent-child associations in health. Panel A holds the rank-rank slopes, the intercepts, and the expected child ranks at various percentiles of interest of the parents (p25 and p75). Panel B shows the intergenerational correlations, where individuals are not ranked but where the estimate is standardized according to equation (2).

For visual purposes, panel A of Table 3 is also shown in a figure: Figure 2 shows binned scatter plots of the relation between parent and child rank for health. All parent-child configurations exhibit approximate linearity which is convenient since only two parameters, the slopes and intercepts from Table 3 Panel A, are necessary to characterize the association between generations.

Table 3 shows that rank-rank slopes for health are in the range 0.11-0.15. Thus, for an increase in parent health of 10 percentiles, child health is predicted to increase by 1.1-1.5 percentiles depending on the particular combination of parent and child. The intercepts are seen to be between 43 and 45. Given linearity this implies that a child born with a parent in the absolute bottom (top) of the health distribution is expected to approximately end up at percentile 43-45 (55-57) herself. Panel A of Table 3 also shows the expected ranks at the 25th and the 75th percentiles of the parental health rank. These are seen to be relatively similar across parent-child combinations. At the 25th percentile they are between 47-48 and the expected ranks are between 53-55 for the 75th percentile. Hence, two children with quite different “starting points”, i.e., a parent 50 percentiles apart in health, are expected to end up relatively close, namely within 6-8 percentiles of each other. In general, a parent in comparably bad health does not imply that the child ends up in the lower percentiles herself. Conversely, the offspring of a parent in relatively good health (say, at the 75th percentile) is expected to end up only at slightly above median health.\footnote{The mobility estimates from Table 3, Panel A can also be interpreted in terms of the correlation between the predicted health rank and objective health outcomes presented in Appendix A. Consider the (statistical) effect on daughter disability rate of “moving” a mother 100 health percentiles with a rank-rank slope of 0.14: in Figure A.1 (b), the daughter disability rate spans 15 percentage points. Hence, the average difference between two daughters with these maximally different mothers amount to a $100 \times 0.14 \times 0.15 = 2.1$ percentage points.}

Considering the rank-rank slopes, it is clear that the strongest transmission
Figure 2: Binned Scatter Plots for Health

Notes: Estimations and samples are identical to Panel A of Table 3. Additionally, the graph shows a non-parametric plot of the conditional expectation of child health rank given parent health rank. The points represent binned averages of child health rank conditional on parental health rank.
seems to be the combinations where the parent is the mother and in particular between mother and daughter. Panel B shows very similar estimates so in this case rank-rank slopes and intergenerational correlations yield identical results (in general, these turn out to be quite similar which is why I report only one of them in most sensitivity checks).

Rank-rank slopes (and correlations) in the range 0.11-0.15 is lower than recent findings for the US (Halliday et al. (2020, 2021)), and lower than what Graeber (2018) finds for Germany but on par with findings from Sweden (Björkegren et al. (2019)) who find rank-rank slopes in the range 0.13-0.15. It is worth noting that the included countries rank similar for health mobility than what is usually found for income mobility (Corak (2013)). Hence, Scandinavian countries appear more mobile than the US, with Germany in between, both for health and income. Also, health as a domain exhibits more mobility than that of income, i.e., lower rank-rank slopes and correlations (see for example Björklund and Jäntti (2012), Halliday et al. (2021) Landersø and Heckman (2017)). Naturally, caveats are necessary here. The above studies of health mobility use different samples, and different empirical approaches in measuring health. Future research should obtain comparable cross-country data and evaluate whether mobility estimates differ when using similar measures, samples and age ranges.

Measuring health is not straightforward and low intergenerational associations may arise for various reasons. However, I obtain similar, or moderately lower intergenerational correlations, sibling correlations (and rank-rank slopes) when experimenting with alternative health measures. In Appendix D, I exclude GP visits and obtain two health measures, one based on principal components, and one based on a “simple fractions” approach.

5.2 Sibling and Twin Correlations

This subsection presents results from the second way of evaluating mobility in (health) outcomes.

---

point difference in disability rate on a median of approximately 2 percent. Mean daughter disability rate is 6.5 percent (not reported), and may be more common to benchmark the 2.1 percentage point difference to. As Figure A.1 (b) shows, the associations between health and the objective outcomes are non-linear, but the simple calculation above provides a flavor of the magnitudes involved.
Table 3: Intergenerational Associations In Health

<table>
<thead>
<tr>
<th></th>
<th>(i)</th>
<th>(ii)</th>
<th>(iii)</th>
<th>(iv)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Son-Father</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Health Rank of Parent ($\beta_R$)</td>
<td>0.112</td>
<td>0.127</td>
<td>0.111</td>
<td>0.145</td>
</tr>
<tr>
<td></td>
<td>(0.005)</td>
<td>(0.005)</td>
<td>(0.005)</td>
<td>(0.005)</td>
</tr>
<tr>
<td>Intercept ($\beta_0$)</td>
<td>45.0</td>
<td>44.1</td>
<td>45.2</td>
<td>43.6</td>
</tr>
<tr>
<td></td>
<td>(0.30)</td>
<td>(0.30)</td>
<td>(0.30)</td>
<td>(0.30)</td>
</tr>
<tr>
<td>Expected rank at p25</td>
<td>47.8</td>
<td>47.3</td>
<td>48.0</td>
<td>47.2</td>
</tr>
<tr>
<td>Expected rank at p75</td>
<td>53.4</td>
<td>53.6</td>
<td>53.5</td>
<td>54.5</td>
</tr>
<tr>
<td>Son-Mother</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Daughter-Father</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Daughter-Mother</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Panel B: Correlations ($\gamma$)

<table>
<thead>
<tr>
<th></th>
<th>(i)</th>
<th>(ii)</th>
<th>(iii)</th>
<th>(iv)</th>
</tr>
</thead>
<tbody>
<tr>
<td>IGC ($\gamma$)</td>
<td>0.113</td>
<td>0.125</td>
<td>0.115</td>
<td>0.142</td>
</tr>
<tr>
<td></td>
<td>(0.005)</td>
<td>(0.005)</td>
<td>(0.005)</td>
<td>(0.005)</td>
</tr>
<tr>
<td>Observations</td>
<td>36,178</td>
<td>36,178</td>
<td>35,282</td>
<td>35,282</td>
</tr>
</tbody>
</table>

Notes: All regressions are performed using Ordinary Least Squares. Robust standard errors clustered on family level are in parentheses.
Figure 3 shows various sibling correlations, ranging from ordinary siblings on the left, to twins in the middle, and to identical twins on the right. Note that twins in the middle panel consist both of monozygotic and dizygotic twins. Recall that non-twin siblings and twins can be observed in the data, but that zygosity is unknown. Thus, the two left panels of the figure compute observed patterns in the data while the right panel relies on certain identifying assumptions explained in the methodology section. The rationale for this whole exercise is to obtain plausible bounds on the total importance of family background, and to evaluate how much of inequality in opportunity in health is captured by traditional parent-child associations such as the rank-rank slope and intergenerational correlation vis-à-vis measuring mobility by sibling correlations.

First, consider the "Brothers" and "Sisters" category, i.e., the (non-twin) sibling correlations. Sibling correlations in health are given as 0.14 for brothers and 0.20 for sisters. In other words, 14 percent of the variance of male health outcomes can be explained by family background while this figure is 20 percent for women. As seen below, this suggests that the intergenerational correlations (and the rank-rank slopes) capture persistence of health quite poorly.

Benchmarking these sibling correlations to other estimates from the literature on health mobility is more difficult than for parent-child associations given that the latter are more frequent. Sibling correlations of 0.14 and 0.20 are lower than what Mazumder (2011) reports for adult self-reported health in the US (0.23 and 0.35). Compared with studies investigating sibling correlations in anthropometric outcomes (see e.g., Mazumder (2008) and Björklund and Jäntti (2012)) these sibling correlations are also low. Similarly, sibling correlations of 0.14 and 0.20 are considerably lower than those found from principal components from self-reported ailments (Halliday and Mazumder (2017)), although here the health measure and the age of the siblings vary considerably from my study. Comparing the sibling correlations to other more in the spirit of adult health outcomes the estimates found here are more on par with sibling correlations in longevity (Piraino et al. (2014)), but again, this comparison is imperfect as the sample of Piraino et al. (2014) is quite different from mine.

Next, consider the remaining (twin) sibling correlations of Figure 3. Unsurprisingly,
Notes: Estimated sibling correlations for both genders based on equation (5). The first block is ordinary siblings, the second block is for twins, the third block is for identical twins (See Table 2). Estimations are based on 36,178 sons from 32,674 families and 35,282 daughters from 31,983 families. Estimates in second and third block for twins are based on 444 male twins, 446 female twins and 430 mixed sex twins. The error bars show 95% confidence intervals from bootstrapped standard errors.
Figure 3 shows that sibling correlations in health increase when their degree of similarity goes up (that is, when moving left to right in the figure, conditional on gender, i.e., moving from siblings to twins to identical twins). Their increasing similarity can be a result of e.g., increased genetic overlap, increased interaction and increased shared exposure to various shocks to the family. The aim here is not to disentangle these forces empirically but to group them collectively as factors beyond the control of the individual and hence to take them into account when gauging equality of opportunity. Unfortunately, it is also seen that the precision of the estimates for monozygotic (i.e., identical twins) is low. In Andersen et al. (2019) I show that employing a different, larger sample leads to more precisely estimated parameters for the health outcome, and the point estimates themselves not changing dramatically. Hence, these twin correlations can be taken at face value with at least some credence. The estimates in the third block of Figure 3 show that male identical twins have a health status correlation of 0.34 and female identical twins have a correlation of 0.38. Hence, the upper bound of the importance of family background for health is at 34-38 percent. Again, this suggest that health is a domain where family background is of less importance than e.g., earnings or education (See Figure 3 in Björklund and Jäntti (2012)).\footnote{The reason is that Björklund and Jäntti (2012) is based on Swedish data and Danish and Swedish mobility estimates generally align.} However, of course, the lack of precision of these estimates should be kept in mind.

5.3 Comparison of Parent-Child Associations with Sibling Correlations

Individuals may experience inequality of opportunity for reasons only poorly reflected in the correlation between their own status and the status of their parents. Table 4 shows the share of the sibling correlation that is captured by the traditional parent-child models according to equation (3). The remaining part is “other factors”. For the parent-child combinations with a son (daughter) as the child, the sibling correlation is that of brothers (sisters). Other comparisons of this type in the literature can be found for income in the US (Solon (1999)) or a variety of labor market outcomes in a Swedish setting (Björklund and Jäntti (2012)).
Table 4 shows that the intergenerational correlations across all parent-child combinations only explain 9-11 percent of the sibling correlations. In the case of women, the highest intergenerational correlation was found to be 0.142 which implies a squared intergenerational correlation of 0.020. As shown in the bottom of Table 4, the parent-child correlation explains \((0.142^2/0.20)\cdot100 = 10.1\) percent of what tends to make sisters similar in terms of health status. For brothers (comparing with the highest intergenerational coefficient, that is, the mother-son association), making a similar calculation yields an estimate of 11 percent. Recalling equation (3) this indicates that the second term, the "other factors", is quite important. In Table 4, “other factors” constitute 89-91 percent of the sibling correlation, or equivalently, the intergenerational correlations only explain 9-11 percent of the sibling correlations.

Here, a comparison to the US can be made. Using an intergenerational health association of 0.23 (Halliday et al. (2021) or 0.30 (Halliday et al. (2020) and a sibling correlation of 0.53 (Halliday and Mazumder (2017)) or 0.25 Mazumder (2011) suggests that parent-child associations explain 10-36 % of the sibling correlation.\(^{13}\) Hence, parent-child associations may miss more of the family persistence in this setting than in the US, in which case interpreting cross-country mobility becomes more difficult. I have also experimented with adding both measures of parental health in the parent-child models which increases explanatory power moderately. However, the share that parental health (in total) explains of the sibling correlation never exceeds 20 percent. It is also not clear whether this is the appropriate comparison as most influential mobility studies using parent-child models often report estimates from pairwise parent-child combinations as in Table 3.

Note also that Table 4 compares results for the (non-twin) sibling correlations. Having instead compared parent-child associations with (identical) twin correlations would have delivered even larger estimates of the importance of “other factors”, or equivalently, that the intergenerational correlations captured the importance of family background to an even lesser extent.

It is in this sense that Björklund and Jäntti (2012) conclude that studying parent-child

\(^{13}\)A caveat here is that the \(\beta\) coefficient is used, not the intergenerational correlation (see e.g., Piraino et al. (2014) and Björklund and Jäntti (2012)).
Table 4: Comparison of intergenerational correlations (IGC) and sibling correlations.

<table>
<thead>
<tr>
<th></th>
<th>(i)</th>
<th>(ii)</th>
<th>(iii)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$\rho_S$</td>
<td>$\text{IGC}^2 (\gamma^2)$</td>
<td>Other Factors</td>
</tr>
<tr>
<td>Son-Father</td>
<td>0.14</td>
<td>0.013</td>
<td>0.127</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(9.1)</td>
<td>(90.9)</td>
</tr>
<tr>
<td>Son-Mother</td>
<td>0.14</td>
<td>0.016</td>
<td>0.124</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(11.2)</td>
<td>(88.8)</td>
</tr>
<tr>
<td>Daughter-Father</td>
<td>0.20</td>
<td>0.013</td>
<td>0.187</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(6.6)</td>
<td>(93.4)</td>
</tr>
<tr>
<td>Daughter-Mother</td>
<td>0.20</td>
<td>0.020</td>
<td>0.180</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(10.1)</td>
<td>(89.9)</td>
</tr>
</tbody>
</table>

Notes: Estimates from Table 3 Panel B and Figure 3. Percent of the sibling correlation attributable to parental health and to other factors is within parentheses according to equation (3). See Table 4 in Björklund and Jäntti (2012) for a similar decomposition with labor market outcomes and anthropometric outcomes.

associations is like studying “the tip of the iceberg” if one wants to understand why status is reproduced in families. They find the “tips of the icebergs” to be between 4 percent (non-cognitive skills) and 43 percent (height) of the sibling correlation. For the US, Solon (1999) reports that forty percent of sibling correlations in earnings are explained by parental earnings.

Hence, my conclusions for health indicate that, relatively speaking, a large fraction of the iceberg is beneath the surface relative to other outcomes, e.g., earnings or income. For example, it may be the case that sibling resemblance in health is the result of similar time preferences or educational levels rather than healthy parents somehow passing on health capital (see e.g., Fuchs (1982)).
6 Discussion

In line with other literature I confirm that parent-child regressions can be seen as the tip of the iceberg when compared to sibling correlations, which is a broader measure of equality of opportunity (Björklund and Jäntti, 2012; Bredtmann and Smith, 2018; Mazumder, 2008; Solon, 1999). As a methodological point this is yet another indication that traditional parent-child associations far from paint the full picture. For health I find that no more than eleven percent of the sibling correlation can be explained by parental health. This may be a hint to policy makers that if they are worried about reproduction of health status in families, parental health is not necessarily the only avenue in which to consider interventions. Rather, it is the broad set of circumstances that are shared in families.

It is beyond the scope of this study to investigate observable mechanisms (or speculate extensively on unobservables), but future research in health mobility needs to consider that the correlation between parental and child health may only be the tip of the iceberg in explaining health reproduction in families. When discussing cross-country differences in health mobility, it is also interesting whether they remain if mobility is evaluated according to sibling correlations rather than parent-child associations.

In line with Björklund and Salvanes (2011) and Björklund and Jäntti (2012) I suggest that the total importance of family background is bounded between sibling correlations and identical twin correlations. Using this approach I find that 14-38 percent of the total variation in health can be explained by family background factors. While this is less than for other relevant outcomes such as income, earnings and educational attainment, it suggests that even in comparably egalitarian societies with universal health insurance, health status is still reproduced within families.

The final conclusion is tentative given the infancy of the literature with its associated difficulties in cross-country comparisons. Nevertheless, it is striking to note that most of the estimates from the literature on mobility in education, income, and also health are consistent with a rule of thumb saying that persistence in social status in Scandinavia is between one half and two thirds of what is found for the US. It is plausible that part of this difference can be attributed to varying levels of income inequality, economic mobility
and availability of universal health insurance versus employer provided insurance.\textsuperscript{14}

This reasoning parallels that of e.g., school quality and income inequality being powerful predictors of commuting zone variation in income mobility (Chetty et al. (2014)), possibly also having a causal effect on mobility levels. This does not imply that countries necessarily would attain higher health mobility rates from one day to the other if they implemented policies to expand health insurance, reduce economic mobility and increase mobility in income and educational attainment; however, it may interpreted as a suggestive pattern that some areas exhibit higher mobility levels than others.

In this study, the latent variable was based on different data than comparable studies (e.g., Halliday et al. (2021)), and was modeled in a simpler fashion (e.g., Halliday et al. (2020)). Future research should obtain comparable cross-country data and use similar health measures, mobility metrics, samples and age restrictions, and evaluate whether mobility estimates vary as a function of institutions, policies and other characteristics, e.g., levels of income inequality. Disentangling the many complex ways in which institutions, policies and societal norms relate to health inequality and health mobility remains an important task for social scientists in the future.

\textsuperscript{14}Fletcher and Jajtner (2019) report lower rank-rank slopes (i.e., higher mobility) for children whose parents have health insurance. Of course, having health insurance is likely not an exogenous variable but at least this finding is suggestive about the role of policies in fostering mobility.
References


Parman, J. et al. (2010). *Gender and intergenerational mobility: Using health outcomes to compare intergenerational mobility across gender and over time* (tech. rep.).


Appendix

A Validation of the health measure

A.1 Validation of main health measure

Below, I test how the health measure relates to events that arguably signal objective health problems. It would be a source of concern a measure of health had no relation to, say, mortality. For parents I show how the health measure predicts mortality. For children the health measure is shown to predict uptake of disability pension and mortality. Figure A.1 (a) shows the 4-year mortality rate for parents across the health distribution split by parent gender. This is the observed mortality for four years after the health observation window, i.e., it is the mortality rate in the years where the parents are between 70 and 74 years old. As the youngest parent birth cohort is 1940 and the data ends in 2014 this is the maximum window in which they can be followed. It is seen that men at the very bottom of the health distribution have a mortality rate of 35 percent, while this figure drops to around 7 percent for men in the very top. For women, there is the same tendency although, as expected, the absolute rates are lower. Thus, the measure does indeed seem to capture some health signal, which is reassuring.

Figure A.1 (b) shows the corresponding graph for the child generation but where the outcome is uptake of disability pension. Here, the health measure is computed based on the years 36-46 as opposed to 36-50 in the main estimation. This is a necessary implementation as the years 46-50 are needed for observing mortality and disability pension outcomes. Disability pension is an outcome with a higher incidence than mortality (for the child generation) and also plausibly serves the purpose of validating the health measure. Figure A.1 (b) shows a clear downward trend in disability rate throughout the health percentiles. For example, women in the bottom of the distribution have a 15 percent chance of taking up disability pension at some point in this four year window, for women at the median it is around 2 percent while it is essentially zero for the most healthy individuals.

Finally, consider the relation between predicted child health and mortality. Figure A.2 shows predicted mortality to be a somewhat decreasing function of health rank, although
Figure A.1: Validation of health measure

(a) Parent Health and Mortality

(b) Child Health and Disability Pension

Notes: The graphs show nonparametric regressions of child and parent outcomes in age interval 46-50 on the respective predicted health ranks. Confidence bands are based on a 95 % level.
Figure A.2: Child Health and Mortality

Notes: Nonparametric regression of child mortality rate in age interval 46-50 on the predicted health rank. Confidence bands are based on a 95% level.

The picture is less clear than what Figure A.1 showed for parents. For males in particular it seems that the signal is strongest in the bottom fifth. As mortality rates for these birth cohorts, 1960-1963, are rather low it is a demanding outcome to predict.

Overall, as expected, being in the lower health percentiles (i.e., being less healthy) is associated with markedly increased probability of being deceased, or on disability pension, over the next four years. Appendix A.2 shows that the above patterns prevail when GP visits are excluded from the health measure. Thus, it is certainly not the case that the main health measure picks up health wrongly due to individuals of high socioeconomic status often seeking doctor advice, and generally, investing more in their health this way. At the very least, if this is happening to some degree it is not driving the results.\(^{15}\)

\(^{15}\)In results not shown I have verified that the relation between objective health outcomes and predicted health ranks are monotone for all income quintiles, although unsurprisingly being at different levels of the health outcome.
A.2 Validation of health measure using only GP visits

Figure A.3, Figure A.4 and Figure A.5 disregard hospital stays and show the relation between objective health events and quintiles of doctor visits. These figures are included to demonstrate that while doctor visits may be thought to capture other aspects than health status (as opposed to hospital stays) it turns out empirically that the least healthy also display the most doctor visits and vice versa. In Figure A.3 it is seen that those fathers with the fewest doctor visits (i.e., those in the first quintile) have a 4-year mortality rate of 7 percent in contrast to 17 percent for those with the most visits (i.e., those in quintile five). For mothers, there is the same gradient, albeit less steep. The figure can be contrasted to Figure A.1 which showed the relation between the health rank extracted from principal components, which included hospital visits and doctor visits. Figure A.1 qualitatively showed the same gradient, although the predictive power in relation to mortality was stronger. Hence, utilising doctor visits and hospital visits in unison gives a more powerful health measure than doctor visits alone, which however exhibits the same relation to objective health outcomes on its own, although to a lesser extent. Figure A.4 shows that for individuals in the offspring generation, uptake of disability pension correlates positively with quintile of doctor visits. For sons the probability of transitioning to disability is virtually zero for the first quintile and nearly five percent for the fifth quintile. In this case, the gradient is even clearer for women (i.e., daughters). This can be contrasted with Figure A.1 which showed similar, but stronger relation to disability when the health measure consisted of both types of health care utilisation. Figure A.5 shows the relation to child mortality and quintile of doctor visits and is primarily included for completeness. As commented upon in relation to Figure A.2, mortality is demanding to predict for individuals around age 50 as it has a very low prevalence. Hence, Figure A.5 shows only modest differences in mortality according to utilisation of doctor visits, most of them statistically insignificant. However, at at the very least, being in the quintile with the highest tendency to visit the doctor is associated with a markedly higher mortality (for sons, around 1.5 percent versus less than 1 percent for the other groups).
Figure A.3: Parent GP visits quintile and Mortality

Notes: Relation between parent mortality rate in age interval 70-74 and GP visits quintile. Confidence bands are based on a 95 % Level.

Figure A.4: Child GP visits quintile and Disability Pension

Notes: Relation between child disability pension rate in age interval 46-50 and GP visits quintile. Confidence bands are based on a 95 % Level.
Figure A.5: Child GP visits quintile and Mortality

Notes: Nonparametric regression of child mortality rate in age interval 46-50 on GP visits quintile. Confidence bands are based on a 95 % level.

B Properties of the Health Measure

B.1 Scree Plots

Figure B.1 shows the scree plots from the estimation procedure by generation and sex. The plots show the eigenvalues on the y-axis as a function of the included factors on the x-axis. The four plots are seen to exhibit a similar pattern. According to two rules of thumb, the Kaiser rule, and the scree test (Brown (2014)), the number of factors to include should be guided by the number of factors with eigenvalue above 1 and where on the graph the curve seems to be genuinely levelling off.\textsuperscript{16} In light of these it seems reasonable to retain only the first component although this does not strictly adhere to the first rule of thumb. However, in all four graphs the first component is seen to, by far, contain the most information, where the added gain from each additional component is very marginal.\textsuperscript{17} Also, the whole

\textsuperscript{16}Informally referred to as “the elbow point”.

\textsuperscript{17}The first component explains 13-14 percent of the overall variability depending on the type of individual, with all component loadings being significant at conventional levels (results available upon request). This explained variability is probably on the low side relative to other applications of this method.
Figure B.1: Scree Plots

(a) Sons

(b) Daughters

(c) Fathers

(d) Mothers

Notes: The graphs show the eigenvalues as a function of the component number.

purpose is to come up with a one-dimensional health measure, which is comparable to e.g., income measures, so in a sense the scree plots can be seen as confirmatory rather than exploratory.

B.2 Factor Loadings

Figure B.2 shows the component loadings for the various types. The interpretation of the component loadings is not straightforward as all the input variables are binary. However, the loadings can be evaluated according to their sign and the relative magnitudes. First, it
is seen that the loadings are relatively similar for sons, daughters, fathers and mothers. Also, some ailments tend to have high loadings, such as cardiovascular diseases and endocrine diseases (e.g., diabetes). Less salient is the loading of varicose veins. When thinking about health in terms of lifestyle diseases and behaviour such as lack of exercise and unhealthy diet, these patterns are meaningful. Unsurprisingly, it is also seen that being in the upper two fifths of the doctor visits distribution has loadings similar to the hospital ailments while the loadings on having less than average doctor visits are negative (i.e., signs of being healthy).

**Figure B.2: Component Loadings**

![Component Loadings](image)

**Notes:** Component loadings from the principal components procedure by generation and sex.

### C Lifecycle Bias

In *Andersen et al. (2019)* I test lifecycle bias in estimations by measuring health at different ages. Here, I supplement with testing whether results change when moving measurement windows of the two generations closer to each other.
A concern is whether the health of parents and children are measured too far from each other in the main specification. In the main specification, child health is measured at age 36-50 and parental health at age 60-70. In the following I perform health measurements closer in time between parents and children, and show the gradual results for rank-rank slopes from moving measurements closer in time. For this purpose I select birth cohort 1960 for whom it is possible to measure health up till age 54. For parents I limit health measurements to age 60-65. In order to still avoid considerable attenuation bias I use measurement windows with a length of five years, and thus, the “closest” measurements this exercise provides are health based on age 49-54 for the child generation, and age 60-65 for parents. Results are shown in Table C.1 and Figure C.1. Since both composition and measurement window length are involved in this robustness check, Table C.1 establishes base results for this sub-sample with altered measurement window.

Figure C.1: Rank-Rank slopes for different child health measurement windows - birth cohort 1960

Notes: The figure shows estimated rank-rank slopes for different child health measurement windows for birth cohort 1960, and where parental health is measured at age 60-65.
Table C.1: Intergenerational Associations In Health - Rank-rank slopes Cohort 1960

<table>
<thead>
<tr>
<th></th>
<th>(i)</th>
<th>(ii)</th>
<th>(iii)</th>
<th>(iv)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Son-Father</td>
<td>Son-Mother</td>
<td>Daughter-Father</td>
<td>Daughter-Mother</td>
</tr>
<tr>
<td>Panel A: Child 36-50 and parent 60-70</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Health Rank Father</td>
<td>0.110</td>
<td>0.119</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(0.012)</td>
<td>(0.012)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Health Rank Mother</td>
<td>0.108</td>
<td>0.133</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(0.012)</td>
<td>(0.012)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Observations</td>
<td>8,731</td>
<td>8,731</td>
<td>8,658</td>
<td>8,658</td>
</tr>
</tbody>
</table>

Panel B: Child 36-50 and parent 60-65

<table>
<thead>
<tr>
<th></th>
<th>(i)</th>
<th>(ii)</th>
<th>(iii)</th>
<th>(iv)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Son-Father</td>
<td>Son-Mother</td>
<td>Daughter-Father</td>
<td>Daughter-Mother</td>
</tr>
<tr>
<td>Health Rank Father</td>
<td>0.109</td>
<td>0.107</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(0.012)</td>
<td>(0.012)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Health Rank Mother</td>
<td>0.105</td>
<td>0.141</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(0.012)</td>
<td>(0.012)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Observations</td>
<td>8,731</td>
<td>8,731</td>
<td>8,658</td>
<td>8,658</td>
</tr>
</tbody>
</table>

Notes: All regressions are performed using Ordinary Least Squares. Robust standard errors clustered on family level are in parentheses. In Panel A, child health is measured at age 36-50 and parental health at ages 60-70 (as in the main specification). In Panel B, parental health is measured at age 60-65.
Panel A shows the base specification that mimics the main specification with the only difference being that the sample is limited to child birth cohort 1960. Panel B replicates Panel A, but where parental health is limited to age 60-65. Panel A yields results close to the main specification in Table 3, which is to be expected as there is no reason to suspect that birth cohort 1960 is fundamentally different from birth cohorts 1961-1963. Unsurprisingly, precision of the estimates drop (standard errors up from 0.005 to 0.012). Moving to Panel B it is seen that three out of four estimates drop, while the ”Daughter-Mother” rank-rank slope increases from 0.133 to 0.141. In general, the differences are small. The decreased measurement length of parental health would tend to decrease estimates, while moving measurement window closer in time to the children’s would possibly increase estimates. However, I can now evaluate the effect of gradually moving child measurements closer to a fixed parent measurement of 60-65. For convenience, Figure C.1 only plots point estimates, but standard errors from these estimations are on par with the ones reported in Table C.1. It is clear from Figure C.1 that there is no clear trend in the estimated rank-rank slopes. If the results in the main specification were an artefact of measurements of children and parents taking place at very different points in the lifecycle, Figure C.1 should show an increasing trend. In other words, for a fixed sample and measurement window length, associations should increase in strength as measurements were performed closer to each other. Instead, estimates from child age 36-41 and 49-54 are strikingly close to each other, although there is small variation in the middle of the interval. In conclusion, the results from the sensitivity checks in this subsection suggests that the exact point in time of health measurements, thereby also varying the distance between parent and child measurements, do not substantially alter the estimated rank-rank slopes.

D Results for alternative health measures

This subsection tests whether the estimated associations change when employing alternative health measures. In particular, I test how the exclusion of GP visits affects results in different specifications, both in a principal components framework and one using “simple fractions”. Recall that the main health measure consists of hospital care indicators and
a number of GP visits indicators. As discussed in the health measure section it may be argued that especially the GP visit component partly reflects some health behaviour, not solely reflecting true health (needs). For example, Fadlon and Nielsen (2019) finds that individuals increase health care consumption (e.g., GP visits) immediately after the occurrence of a health shock to someone in the family, or a close coworker. Doorslaer et al. (2004) show that for a sample of European countries, including Denmark, there is income-related inequality in number of GP visits. However, all Danish income-related inequality in number of GP visits disappears when controlling for health needs. In reality, GP visits probably reflects both objective health levels (recall Figures A.3, A.4 and A.5), and some component of health care behaviour, whether interpreted as risk aversion, or some other factors. The question is, however, whether inclusion of GP visits in an aggregate health measure, as largely revealing something about "true" health status is misleading or not.

Table D.1 shows the results from parent-child correlations and sibling correlations from employing the alternative health measure. In Panel A, in contrast to the main specification, principal components is not used, but instead each individual is assigned a simple fraction counting the number of hospital categories from Table 1 the individual has been associated with over the observation period divided by the number of categories (14). This measure is arguably more transparent than a health measure based on principal components. The health measure of Panel B of Table D.1 has exactly the same input variables but employs the principal components procedure also used in the main specification. What distinguishes the two measures from the main measure is that they exclude GP visits. The measure in Panel A is similar to the sensitivity check of Halliday et al. (2021), where they show how a simple "fractions measure" of objective ailments provide similar results as time-averaged self-reported health status. For brevity, Table D.1 excludes the twin estimations. Moreover, Table D.1 solely reports correlations, not rank-rank slopes, as the measure in Panel A is far from being continuous. Consequently, the results in Table D.1 are directly comparable to Panel B of Table 3, which also showed the intergenerational correlations, and the first column of Figure 3 which showed the (non-twin) sibling correlations.

In general, Panel A and Panel B of Table D.1 show very similar results. Compared to Table 3, the parent-child correlations are moderately lower (in the range 0.08-0.11
versus in the range 0.11-0.15). Similarly, compared to Figure 3 the sibling correlations are moderately lower (in the range 0.10-0.15 versus in the range 0.14-0.20). Whether one believes that GP visits belongs in a measure of health status, or that GP visits solely reflects health behaviour, it is probably not surprising that the estimated associations slightly decrease when leaving it out of the health measure. In that sense, the broad conclusion that health persistence is relatively low, and decisively lower than comparable studies find for the US (Halliday et al. (2020, 2021)) and Germany Graeber (2018)), still stands.

Even though GP visits is the component one might suspect not purely reflects true health, a case may also be made for the hospital indicators. For Denmark, Doorslaer et al. (2004) finds that there is income-related inequity in specialist visits, that is, that the upper parts of the income distribution exhibits higher utilisation of specialist services, conditional on health needs. However, for Denmark this is driven by a higher number of visits, conditional on going at least once, whereas the probability of going to the specialist at least once does not display income-related inequity. The variables in the main specification precisely enter as binary variables, not counting the number of visits (Table 1).
Table D.1: Alternative health measures - excluding GP visits

<table>
<thead>
<tr>
<th></th>
<th>Intergenerational Correlations</th>
<th>Sibling correlations</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(i)</td>
<td>(ii)</td>
</tr>
<tr>
<td>Son-Father</td>
<td>0.0798</td>
<td>0.0814</td>
</tr>
<tr>
<td></td>
<td>(0.005)</td>
<td>(0.005)</td>
</tr>
<tr>
<td>Son-Mother</td>
<td>0.0978</td>
<td>0.1069</td>
</tr>
<tr>
<td></td>
<td>(0.005)</td>
<td>(0.005)</td>
</tr>
<tr>
<td>Daughter-Father</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Daughter-Mother</td>
<td>0.113</td>
<td>0.141</td>
</tr>
<tr>
<td></td>
<td>(0.016)</td>
<td>(0.016)</td>
</tr>
<tr>
<td>Brothers</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sisters</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Panel A: Simple fractions

<table>
<thead>
<tr>
<th></th>
<th>Health Father</th>
<th>Health Mother</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0.0798</td>
<td>0.0978</td>
</tr>
<tr>
<td></td>
<td>(0.005)</td>
<td>(0.005)</td>
</tr>
<tr>
<td></td>
<td>0.0814</td>
<td>0.1069</td>
</tr>
<tr>
<td></td>
<td>(0.005)</td>
<td>(0.005)</td>
</tr>
</tbody>
</table>

Sibling correlation

|                    | 0.113         | 0.141         |
|                    | (0.016)       | (0.016)       |

Panel B: Principal components

<table>
<thead>
<tr>
<th></th>
<th>Health Father</th>
<th>Health Mother</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0.0775</td>
<td>0.0824</td>
</tr>
<tr>
<td></td>
<td>(0.005)</td>
<td>(0.005)</td>
</tr>
<tr>
<td></td>
<td>0.0885</td>
<td>0.0990</td>
</tr>
<tr>
<td></td>
<td>(0.005)</td>
<td>(0.005)</td>
</tr>
</tbody>
</table>

Sibling correlation

|                    | 0.104         | 0.1539        |
|                    | (0.016)       | (0.016)       |


Notes: The table shows results from specifications where GP visits are left out. Panel A uses the fraction of hospital conditions the individual has been associated with over the observation window, whereas Panel B uses these indicators in a principal components procedure as in the main specification. Otherwise, estimations are similar to the results presented in Table 3 and Figure 3.
References


Parman, J. et al. (2010). *Gender and intergenerational mobility: Using health outcomes to compare intergenerational mobility across gender and over time* (tech. rep.).


