Childhood Illness and Socioeconomic Outcomes in London, 1870 – 1911*

Krzysztof Karbownik† Anthony Wray‡

July 31, 2013

Abstract

We study the long-run effects of childhood illness on socioeconomic outcomes in London, England, between 1870 and 1911. Poor childhood health may have restricted social mobility and contributed to income inequality. We match patients hospitalized during childhood to censuses of England to identify siblings residing in the same household, who would have experienced similar living standards during childhood, and link all individuals to the 1911 census to observe outcomes as adults. We use sibling-fixed-effects models with identifying variation from heterogeneity in disease incidence, severity and timing. Preliminary results suggest that hospitalized individuals were 5 to 10 percent less likely to be married in 1911 than their siblings. Results are robust to varying restrictions on the quality of matched observations in the sample, and the differences in ages between patients and their siblings. We also find some evidence that childhood illness was associated with lower occupational wages and physical skill requirements.

*We are grateful to Susan Hawkins for providing us access to data from the Historical Hospital Admission Records Project (www.hharp.org). Karbownik appreciates the support of a Hedelius Scholarship that permitted him to conduct this research while visiting Northwestern University. Wray acknowledges the support of the Robert Eisner fund and a Northwestern University Graduate Research Grant which permitted him to travel and conduct research for this project. We would like to thank Joe Ferrie and Joel Mokyr for very helpful comments. All errors of omission are our own.

†Department of Economics, Uppsala University, Kyrkogardsgatan 10B, 751 20 Uppsala, Sweden; Phone: +1 224 235 5870; E-mail: Krzysztof.Karbownik@nek.uu.se.

‡Corresponding author; Department of Economics, Northwestern University, 2001 Sheridan Rd., Evanston IL, 60208, USA; Phone: +1 224 241 4245; Email: awray@u.northwestern.edu.

1 Introduction

This paper studies the long-run effect of poor childhood health on socioeconomic outcomes – including occupational wages and skill requirements – in London, England, between 1870 and 1911. It addresses a series of open questions in the literature on the long-term consequences of the early childhood environment. How much of differences in adult income and human capital accumulation between siblings can be accounted for by childhood illness? To what extent does illness experienced before the age of five matter more than a case of illness experienced later in childhood?

Despite extensive evidence from different countries that in utero health insults and low birth weight can have significant long-run consequences (Black et al., 2007b; Figlio et al., 2013; Oreopoulos et al., 2008; Royer, 2009), much remains to be learned about how health shocks during childhood affect long-run socioeconomic outcomes. Recent interdisciplinary studies have identified epigenetic variation, or the regulation of gene expression, as a potential biological mechanism for explaining the persistent effect of the early-life environment on health outcomes. They have found correlations between early-life socioeconomic conditions, epigenetic modifications, and future health outcomes, but have lacked causal evidence (Lam et al., 2012; Miller et al., 2009). While the fetal origins literature (Barker, 1995) has relied on the sharp timing of shocks for identification (Almond, 2006), we know relatively little about the impact of health shocks that have a longer exposure window or insult individuals repeatedly. Moreover, little empirical evidence has been presented about the relative impact of health shocks at different periods of childhood (Almond and Currie, 2011).

This study introduces historical hospital admission registers as a previously unexplored source of individual-level data for studying the impact of childhood health shocks on adult socioeconomic status. Historical evidence is valuable because it allows the researcher to follow individuals over their life-course.[1] We construct a dataset consisting of hospitalized patients, and their closely spaced siblings, who lived in the same household during childhood.

[1] Currie and Moretti (2005) and Lindahl et al. (2012) use historical data to measure the intergenerational transmission of socioeconomic status.

The empirical strategy relies on sibling fixed effects to gauge the causal effect of childhood disease on occupational skills and demographic outcomes, such as whether an individual is married, has children, or is listed as the household head in the 1911 census of England. Siblings form a natural control group for the patients, despite differences in health and genetic variation, as they would have experienced similar living standards and housing conditions during childhood.[2] We link the hospital admission registers to the population census records of England, to obtain long-run occupational outcomes.[3]

Our project offers a series of improvements over Currie et al. (2010), which asks similar research questions and employs a similar sibling-fixed-effects estimation framework. That study uses the public health insurance records of children in Manitoba, Canada, to show that poor child health is associated with greater welfare participation in young adulthood, primarily because it predicts poor health later in adolescence. It shows that while recurring health problems alone affect high school enrollment, the diagnosis of mental health problems before age 3 – particularly for attention deficit hyperactivity disorder (ADHD) – is predictive of poorer outcomes regardless of whether health problems occur later in childhood. We extend these findings by tracking individuals over the course of their working careers (up to the age of 41) and measuring occupational outcomes, which Currie et al. (2010) do not observe. Although they measure health status throughout the entirety of childhood, the majority of health problems are limited to injuries, asthma and mental health problems. In comparison to a present-day developed country, the disease environment of late-19th century London contained much richer variation in the set of diseases, which had much larger effects on childhood health. This variation allows us to identify heterogeneous effects of childhood health by disease type. While these results are not applicable to a developed-world context, they may be relevant to large cities in the developing world today. Despite these advantages, our study does face some data limitations. In particular, we lack data on childhood health

[2] This strategy gets us as close as possible to the genetic similarity found in a sample of dizygotic (fraternal) twins, the most common variation of twinning. Two confounding factors that emerge in the case of non-twin siblings are different in utero environment and gestation length.

[3] In the future, we plan to transcribe the father's occupation from the census manuscripts to examine whether income can mediate the link between childhood illness and long-run outcomes – another issue that the literature has yet to resolve.

status for the entire population – in our case, London, England – and we must deal with the measurement error in transcribed historical data.[4]

A challenge for this project is to provide evidence on the mechanism underlying the link between health and income observed in the data. Poor health and disease can depress individual productivity by impeding the acquisition of human capital and limiting educational attainment (Bleakley, 2007, 2010). Childhood illness can impair physiological and cognitive development, which in turn can lower the quality of education received and the productivity of human capital acquired (Nilsson, 2009; Wolpaw Reyes, 2012). The effect of illness can work through not only biological changes, but also the inability to attend school or enter into an apprenticeship.[5] Childhood health also has an indirect effect on income by changing the incentives for further investment in human capital and the allocation of resources across children (Parman, 2012). Birth order or the gender composition of children in a household may determine whether parental investment reinforces or compensates for health shocks (Rosenzweig and Zhang, 2009). Furthermore, childhood illness imposes an externality on others, particularly classmates and siblings, through the transmission of infections. A correlation between child health and adult earnings can also arise if illness in childhood persists and has a lasting effect on adult health (Case et al., 2005).

This project also contributes to the standard-of-living debate and the study of the welfare costs associated with the urban penalty in late-Victorian England (Szreter, 2005; Wohl, 1983). The overcrowded housing and poor sanitary conditions of working class neighborhoods, from which hospital patients were primarily drawn, weakened the resistance of sufferers to secondary infections and made them susceptible to spreading diseases within the household and at the workplace. Childhood illness could leave an individual too weak to complete apprenticeships or perform physically demanding tasks (Horrell et al., 2001). Past illness that resulted in lower cognitive ability may have kept individuals from adopting pre-

[4]In Section 2.3, we address concerns about sample selection by comparing descriptive statistics for our matched sample to the population. In Section 3.2, we discuss how measurement error may affect the interpretation of our estimates.

[5]In the future we plan to use the likelihood of school enrollment by age categories for patients relative to their healthy siblings to evaluate the importance of the human capital accumulation channel.

ventative health measures or entering skilled occupations, forcing households into a vicious cycle of illness and poverty. Thus, poor childhood health may have restricted social mobility and contributed to income inequality in early-20th century England.

2 Data and institutions

2.1 Data

We have obtained the records of children hospitalized in late-19th-century London at one of three children's hospitals: the Hospital for Sick Children at Great Ormond Street, Cromwell House Convalescent Home, and Evelina Hospital (Kingston University, 2010). These records have been transcribed from inpatient admission registers as part of the Historical Hospital Admissions Research Project (HHARP). The sample contains 16,461 admissions by male patients born between 1870 and 1890, and hospitalized before the age of 13. We hand-coded unique patient identifiers and grouped the admissions into 11,616 "patients" (see the Appendix for details). An individual entry typically includes the patient's name, age in years, and residential address; a description of the patient's disease or cause of admission; dates of admission and discharge; and a categorical variable for the outcome (cured, relieved, not relieved, or died).[6] As far as we know, this source has not been exploited previously by economic historians.

We are extending our study of childhood illness by incorporating available records of in-patient admissions to general hospitals and infectious disease hospitals in London. We have transcribed the records of over 40,000 inpatient admissions to St. Bartholomew's (Barts) Hospital in London between 1874 and 1906 for children aged 0 to 18 when admitted. In combination with the children's hospital admissions from the HHARP database, our transcribed records account for about 15 percent of the London hospital market in a given year. We have also digitized the inpatient admission records from a handful of other infectious disease

[6] The most common diseases for which patients were admitted to the hospital included bronchitis or pneumonia, diseases of the hip, tuberculosis, chorea, empyema, diphtheria, rickets, carries spine, rheumatism and diseases of the knee.

and general hospitals in London (Guy's, London Fever, Royal Brompton, Royal London, and St. Thomas'). In future work, we can extend our annual coverage to 45 percent of the London hospital market with the transcription of these records.[7] As the process of cleaning and linking the Barts Hospital data is ongoing, the results reported in this paper are based solely on the HHARP children's hospital data.

The primary outcome of interest is individual socioeconomic status in 1911, which is captured by rankings of occupational titles reported in the census. We use the occupational measures in England and Kilbourne (1988) from the 1980 US Dictionary of Occupational Titles, which are the best available proxies for human capital accumulation. The richness of this dataset allows us to determine whether the results are driven by occupational income, prestige, or physical or cognitive skill requirements. We will also consider alternative measures of socioeconomic status given that the distribution of occupational skill requirements likely changed substantially between 1911 and 1980 due to technological improvements. We obtain the demographic outcomes from the 1911 census of England.

There are several novel features of the linked hospital admissions and census data which provide a means to improve upon the existing literature that addresses the effect of childhood health on long-run outcomes. Most importantly, the data include not only the childhood illnesses suffered, but also adult occupational outcomes. The inpatient registers contain the timing and duration of illness, which are useful for identifying critical periods of health during childhood and the effect of illness severity. The registers also include multiple admissions for many individuals and allow us to examine the effect of age-specific illness for a wide range of diseases. In particular, we can distinguish between plausibly exogenous infectious diseases and injuries, on one hand, and chronic conditions, on the other. The measures of health constructed from the historical hospital data are more informative and reliable than the self-reported health measures used in contemporary studies (Smith, 2009), or the proxy measure of adult height used in historical studies (Parman, 2010).

[7]This is a rough estimate based on the annual number of inpatient admissions to the hospitals in our data and a total of 68,319 inpatient admissions in London in 1893 (Waddington, 2000, p. 29). We have not located an estimate of the total number of children admitted to London hospitals anually.

2.2 Historical Background

A potential concern for the external validity of our estimates is that a selected set of individuals were admitted to hospitals in late-19th century London. Voluntary hospitals were established with the intention of providing free care to the "deserving" poor: the class of patients above pauperism, but below the capacity of paying for medical care (Lomax, 1996, p. 41). Hospitals sought to deny admission to patients in receipt of public aid under the Poor Law, but it is unclear whether they could distinguish between the needy poor and those on relief. During the 1860s and 1870s, the patients of voluntary general hospitals were drawn primarily from the working class. General practitioners had emerged as the doctors of the middle class, treating patients in the privacy of their homes. As hospital budgets fell into deficit in the 1870s, a belief emerged among hospital staff that patients who could afford the fees for private medical service were abusing the hospital system (Lomax, 1996, p. 9). Efforts by hospital management to identify and restrict access of patients who could afford the payment of admission fees proved to be unpopular and, ultimately, unenforceable. Hospitals came to recognize that, whereas the wealthy had no interest in entering an institution, middle class patients were underserved by hospital care. Furthermore, the growing knowledge of the infection process made it clear that hospitals ought to provide care to all classes of society (Abel-Smith, 1964, p. 119). London general hospitals began accepting payments from patients in the 1880s, as the cost of treating inpatients rose and fundraising sources dried up.

In practice, doctors had considerable authority over the types of cases that were admitted from the pool of outpatients that far exceeded a limited supply of hospital beds. Medical staff would selectively admit patients from the outpatient department, favoring medically "interesting" cases and acutely sick patients in order to demonstrate immediate results for teaching purposes (Abel-Smith, 1964, p. 39). Doctors were opposed to admitting chronic or incurable cases in order to avoid having a bed occupied for a lengthy period of time, or to limit the number of deaths at the hospitals (Waddington, 2000, p. 9). The more patients a hospital treated and the fewer deaths reported, the more attractive the hospital appeared to donors when it applied for funds. Thus, the set of diseases admitted to hospitals cannot be considered representative of the population in light of the demand and supply side selection processes.

A different set of considerations influenced the hospital admission and discharge decisions for children. Until the mid-19th century, it was a generally accepted belief that infants should not be admitted to hospitals, so as to avoid the presumed ill-effects of separating the baby from the mother. However, as the reputation and quality of children's hospitals improved, parents and doctors began to view the hospital as the most suitable environment for acutely sick infants. Although doctors could only provide limited medical treatment of the underlying causes of disease, hospitals offered bed rest, nursing care and a more adequate diet than many children received at home. In the 1870s, children's hospitals also began providing isolation facilities for patients with infectious diseases which limited the spread of infection within the hospital (Lomax, 1996). Doctors often extended the hospital stay of patients when conditions at home were unsatisfactory. In the case of older children, parents may have been reluctant to admit a sick child who contributed to the family's income. Tanner (2007) provides anecdotal evidence that parents removed their children from the hospital against the doctor's wishes when they were needed to perform household chores or look after siblings. Thus, the length of stay in a hospital may reflect the constraints on hospital bed capacity and parental budgets as much as the severity of a disease. As a result, we observe length of stay, conditional on a given severity of illness, with measurement error.

2.3 Summary Statistics

In an ideal setting we would observe the childhood health status of the entire population, but we are limited to observing the health status of children who appear in the hospital records. In Table 1, we address concerns about selection by presenting household-level mean characteristics for the sample of hospital patients matched to a census of England (either 1881, 1891 or 1901), as well as all children in London and England who were enumerated in the 1881 census. When we restrict attention to households in London with at least two male siblings and the mother present, we find that households with hospital patients are similar to those without patients on a number of observable dimensions: uniqueness of names, total number of (male) children, the presence of the parents in the household, the age at first birth of the mother, and the labor force participation rates of the mother and father. The one noticeable difference is that siblings in households with hospital patients are on average 1.7 years younger (8.1 versus 6.4 years) than siblings in other households in 1881. The average age of siblings in the pooled sample is higher than the 1881 sample since the median year of admission in our sample is 1887, so we observe patients in 1891 and 1901 when they are older. The distribution of occupations held by fathers with hospitalized children is less dispersed. Fathers in households with hospital patients are less likely to work in high socioeconomic status (professional and clerical) and low socioeconomic status (unskilled) occupation, but are more likely to work in skilled and low-skilled occupations. This finding is consistent with the tendency of middle-class families to receive medical care from private doctors and hospitals to limit the admission of paupers. Overall, we conclude that the hospital patients in our sample are not drastically different from a representative household in the population with at least two male children.

3 Matching Procedures

We locate the childhood household of residence by matching all patients to each of the 1881, 1891 and 1901 censuses of England (we refer to this as a Stage 1 census match). We use the complete transcription of the 1881 census from the North American Population Project (NAPP), as well as the transcriptions of the 1891 and 1901 censuses on a genealogy website (Minnesota Population Center, 2008; Schürer and Woollard, 2003). We identify male siblings based on the names of parents reported in the census and the organization of records into households on the genealogy website.

We search for patients in the censuses on the genealogy website using first and last name, a 10-year band (+/- five years) around the median reported birth year across all admissions for a patient, and the county of residence.[8] After obtaining the search results, we calculate the Levenshtein[9] string distance between the first names (last names) reported in the HHARP records and the census, as well as the distance between the birth years imputed for the two sources. We restrict attention to records with string distances in names no greater than two units, and birth year differences no greater than two. We require an exact match for county of residence since there are no transcription errors in this field.

We use a combination of the iterative matching strategy used prominently in the literature (Abramitzky et al., 2012; Long and Ferrie, 2013), and the probabilistic matching approach implemented by Mill and Stein (2012) to determine successful matches. The iterative matching strategy arbitrarily prioritizes year of birth as a matching criteria, as it only considers records with an exact match for year of birth in the first stage of the algorithm. Such a strategy is not appropriate in our context since we do not observe the exact year of birth in either of our sources.[10]

Instead, we begin by ranking potential matches using the sum of the string distances for first and last names, and the difference in imputed birth years ("match distance"). For patients residing in the county of London, we add a one unit penalty to the match distance for census observations which resided in a registration district different from that reported in the HHARP records.[11] We add two iterative steps to resolve cases with equivalent match distances:

1. If multiple census records have the same match distance, we only keep those with the smallest combined string distance for first and last names.

[8] The +/- 5-year band around the median birth year was chosen to avoid obtaining duplicate results across multiple searches. The rule captures all potential matches for each reported birth year, since we allow a matched observation to have up to a +/- 2-year error in birth year.

[9] The Levenshtein string distance algorithm does not account for the length of the string, which results in poor matches for shorter strings. An alternative would be to use the Jaro-Winkler distance which adjusts the quality of a match if the first three letters of a string match.

[10] The hospital records report the age in years on admission, while the census reports the age at the most recent birthday.

[11] Residential district is only reported in the hospital records for patients residing in London. We plan to identify the registration district based on the reported street address using street indexes from the census.

2. If any remaining records in a set of multiple matches have the same registration district, we drop records that do not match on registration district.

Our empirical analysis is primarily based on the set of patients with a unique match to a census record that minimizes the match distance between the two sources. We also consider a more restrictive matched sample which permits a maximum discrepancy of one unit for each name strings and birth year, and requires that there are no other potential records within one match distance unit of the matched record.

In Stage 2 of the matching procedure, we search for individuals from the Stage 1 sample in the 1911 census of England to obtain socioeconomic outcome measures. The patients were of working age (21 to 40) when enumerated in 1911. We use the county and parish of birth obtained from the Stage 1 census match as search criteria, in place of residential location. Due to a high frequency of transcription errors in the 1911 census, we allow for differences in the parish name strings up to a distance of two.[12] We follow the procedures for string distance tolerance and match quality ranking described above. The final samples consist of all households with at least two male siblings, including at least one male patient, matched to a census during childhood and the 1911 census.

4 Empirical Framework

4.1 Main Specifications

Our empirical framework largely follows what has been standard in the literature. We estimate sibling fixed effects models in which the regressors of interests are indicators for admission to a hospital by sibling i from household j. We estimate the effects of childhood health insults on the occupational skill requirements and wages that we assign to the occu-

[12]For each name and birth year combination, we perform at least two searches with either "County, England" or "Parish" as the birth place parameter (for individuals born in London, we also perform a search with "*London*" entered in the birth place field). The searches will collect all cases where the parish name is spelled correctly, regardless of whether the county name is also included. The first case will also capture transcription errors in the parish name, if the county name is also provided. Searches with wildcard characters in the parish name were not effective.

pational titles reported in the 1911 Census of England. In particular we estimate following equation:

$$Occupation_{ij}^{1911} = \beta Hospitalized_{ij} + \gamma X_i + \alpha_j + \varepsilon_{ij} \tag{1}$$

where i indexes individuals, j indexes household, and $Occupation_{ij}^{1911}$ is a measure of socioeconomic status in 1911 for individual i from childhood household j (occupation-specific skill requirement, or mean annual occupational income in 1911), X_i is a vector of individual characteristics (age in 1911, age in 1911 squared, and inferred birth order), α_j denotes unobservable determinants of the outcomes that are specific to household, and ε_{ij} is a heteroskedasticity-robust error term that represents sibling-specific unobserved characteristics.[13] We cluster standard errors at the sibling level to account for intra-sibling correlation in variances. Since the outcome variable $Occupation_{ij}^{1911}$ only varies across occupations, the coefficients of interest, β, capture the effect of hospitalization on sorting into occupations with lower skill requirements, or on sorting across occupations with different income levels, but not on success within an occupation. In the preliminary results reported in this paper we also present specifications similar to Currie et al. (2010) in which we substitute the dummy for ever-hospitalized with indicators for hospitalization during one of three age groups: $\beta_1 Hospitalized_{ij,0-3} + \beta_2 Hospitalized_{ij,4-8} + \beta_3 Hospitalized_{ij,9-12}$. We also present specifications with outcomes different from occupations, such as whether an individual is married or is listed as the household head in the census, or the number of children in 1911.

4.2 Identification

Ordinary Least Squares (OLS) estimation of (1) and (2) would produce biased estimates of β if hospitalization was correlated with household conditions such as housing and nutritional quality – in other words, if there were unobservable determinants of occupational choice that were correlated with disease incidence. We estimate a sibling fixed effects model to address

[13]Note that the index j refers to the household in which the individual resided as a child. We do not require siblings to be living in the same household as adults.

the potential bias due to correlation between unobserved household characteristics, health status, and occupational outcomes. We restrict attention to closely spaced siblings, who are likely to have shared the same household conditions. The sibling fixed effect model essentially differences out any household specific confounder and identifies β based on between-sibling variation in birth conditions. Given that we do not observe birth weight or gestational length, we cannot interpret β as the effect of an exogenously induced illness.[14] Our estimates are identified by variation in not only pre-natal health, but also heterogeneity in disease incidence, intensity and timing during childhood. We interpret the hospitalization effect as a deficiency in unobserved health capital, which encompasses congenital conditions traced to birth, domestic living conditions, net nutritional intake, exposure to infectious diseases, and hospital care.

The sibling fixed effects estimation strategy is attractive due to its ability to control for unobserved factors common to the household. However, identification in the model requires the assumption that there are no unobserved sibling-specific factors that are correlated with illness and explain occupational skill requirements. A potentially confounding effect may arise if parents make behavioral adjustments by altering the allocation of resources between siblings in a way that reinforces or compensates for the effect of the health shock. In the case that parents compensate for health shocks, in the interests of equity, the fixed effects method will under-estimate the direct biological impact of the shock, and a significant coefficient can be interpreted as a lower bound. A more problematic case occurs when parents systematically reinforce the effect of the health shock – perhaps motivated by a higher return to investment in the healthy child – and then the true effect cannot be disentangled from the behavioral adjustment. The role of parental investment may be less of a concern in the proposed setup since the sample is restricted to males, which eliminates the effect of gender-based discrimination.

A potential concern for internal validity arises from inferring the health status of siblings

[14] In future specifications, we will restrict attention to the effect of accidental injuries and infectious diseases, which are plausibly exogenous and independent of birth endowment. In the late-19th century, households with higher socioeconomic status were not conferred an advantage in disease prevention due to the imperfect knowledge of proper hygiene practices, as well as the sources and transmission mechanisms of disease.

who do not appear in the hospital records. It is not possible to rule out that the "healthy" siblings also experienced childhood illness. Siblings of hospitalized patients faced a greater than average risk of illness given the likelihood of spillover effects from disease transmission within the household. Parents may have sent their children to a hospital for which the admission registers have not survived, or may only have been able to afford sending a single child to the hospital. These sources of measurement error would bias the estimated coefficient towards zero and a statistically significant negative coefficient must be considered as a lower bound. Another potential concern is that we can only identify the coefficient of interest from households with occupational outcomes observed for both hospitalized and healthy siblings. A correlation between the success of matching an individual to the 1911 census and early childhood health status could thus represent a source of bias.

5 Empirical Evidence

5.1 Descriptive Statistics

We begin with a sample of 5,430 patients and 10,617 healthy siblings pooled across the 1881, 1891 and 1901 censuses of England. In Table 2 we show the number of matched observations of patients and healthy siblings under each of the sample restrictions we impose in our regressions. When we restrict attention to healthy siblings within 5 years of age of individuals hospitalized as children, we match as many as 995 patients and 1,235 siblings, and as few as 389 patients and 499 siblings in our most restrictive sample with only precise and unique matches.[15] In Table 3 we present descriptive statistics for variables reported in the 1911 census, and for the occupational skill requirements from the 1980 dictionary of occupational titles that we assign to the occupational strings in the census. The means are reasonably stable across our samples, although individuals in our most restrictive sample are less likely to be married (45.6 percent versus 55.6 percent) and less likely to be listed as a household head (45.0 percent versus 54.4 percent) in 1911.

[15] See the Appendix section 7.8 for definitions of precise and unique matches.

5.2 Main Results

In Panel A of Tables 4 to 8 we report coefficients on an indicator for whether an individual was ever hospitalized as a child, while in Panel B we report coefficients from an alternative specification in which we include three separate indicators for hospitalization by age groups: 0 to 3, 4 to 8 and 9 to 13.[16] Each specification is an ordinary least squares estimate of a sibling-fixed effects model that includes age in 1911 and its quadratic, and inferred birth order within the household. We cluster standard errors at the sibling level. In column (1) we estimate the model on the complete sample of childhood households for which we matched both a hospital patient and a healthy sibling to the 1911 census. In columns (2) to (5) we restrict attention to healthy siblings within 5 years of age of a hospitalized patient, and present samples with the different restrictions on match quality described above.

We now turn to our main regression results in Tables 4 and 5, which examine the effect of childhood illness on occupational wages and skill requirements. In Table 4, the dependent variable is the log of the mean occupational wage in 1980 US dollars. We find negative coefficients on the indicator for hospitalization between the ages of 0 to 3 in each of the specifications, although they are for the most part estimated imprecisely. The estimated coefficient of -0.050 for the full sample in Column 1 is statistically significant at the 5 percent level.[17] These findings are consistent with an interruption to the cognitive development that occurs during the early childhood period and impacts adult occupational outcomes.

In Table 5 the dependent variables are a series of occupational cognitive and physical skill requirements. In Column (7) we find that hospitalization during childhood is associated with entering occupations that have a lower physical strength requirement, which is not surprising given that late-19th century illness often took a considerable physical toll.[18] Although the majority of our estimates lack statistically significance, we are encouraged by

[16]Each indicator variable is equal to one if an individual appears at least once in the hospital records during the stated age range.

[17]We hesitate to interpret the magnitude of the coefficient given that the units of the dependent variable are 1980 US dollars, while our sample includes individuals observed in 1911 England.

[18]We are puzzled by the positive and statistically significant effect of hospitalization on the eye, hand and foot coordination requirements of an occupation.

the consistency of the negative coefficients on the hospitalization indicators across various specifications. This suggests that the occupational skill requirements from the 1980 dictionary of US occupational titles provide an imperfect proxy for the distribution of occupations in 1911 England, and that we lack sufficient sample size. In future work we plan to supplement these findings by collecting wage data from England around 1911 and categorizing the occupational titles by the HISCLASS ranking of socioeconomic status groups. We are currently transcribing additional hospital records, which will allow us to triple our current sample size and thus reduce standard errors.

5.3 Other Economic Outcomes

We next consider how childhood illness may affect other socio-economic outcomes in addition to occupational skills and wages. Although we currently lack a precise classification of occupational titles, the 1911 census of England reports a series of binary outcomes of interest. In Tables 6 to 8 we explore the impact of childhood illness on the probability of being married or listed as a household head in the 1911 census of England, and the impact on the number of children reported in 1911, respectively.[19] In Column (4) of Table 6 we find that childhood illness is associated with a 5.5 percent reduction in the probability of being married in 1911 (statistically significant at the 10-percent level). This effect appears to be driven by illnesses that occur between the ages of 4 and 8, as hospitalization during this period of childhood is associated with a larger 7.7 percent marriage penalty. This suggests that childhood illness can compromise the desirability of an individual as a marriage partner, or lower the returns to starting a family. Finally, in Tables 7 and 8, we find mostly insignificant effects on the household head indicator and the number of children, although the coefficients are consistently negative.

[19]Since we have only collected information on the names and ages of household members of individuals in our sample, we identify children by the following conditions: individuals in the same household who share the same surname and are at least 20 years younger than the potential father. Individuals with children in our sample are on average 34.3 years of age and have 2.6 children.

5.4 Robustness to Specifications

Up to now we have presented results for specifications that arbitrarily restricted the sample to include only health siblings that were within 5 years of age of a patient in the same childhood household. In Table 9 we show that our results for the effect on marriage probability are robust to restricting the sample to different bandwidths of age differences between patients and healthy siblings.[20] It is interesting to note that column (2) shows a much larger effect of childhood illness on the probability of marriage when we restrict attention to a smaller sample of healthy siblings within 2 years of age of individuals hospitalized as children (relative to the main sample restriction of +/-5 years of birth, which we reproduce in Column (2)). The estimated coefficient not only implies an 11.8 percent reduction in the probability of marriage, but is also more precisely estimated. The closely spaced healthy siblings in the restricted sample are the more appropriate comparison group because they are less likely to differ from patients along unobserved dimensions. As we increase the sample to include siblings spaced further apart, we risk introducing unobserved within-sibling differences due to changes in the living standards of the household over time. As we increase our sample size with newly digitized patient records, we hope to present more results with samples restricted to closely spaced siblings.

5.5 Mechanisms

An important question of interest is what mechanism explains the persistent effect of childhood illness. We have already discussed how education could represent a potential mechanism, as illness could impair a child's ability to learn or attend school. In future work, we will consider school enrollment as an intermediate outcome of interest to examine how poor health can impede human capital accumulation. We will use an indicator for whether an individual appears in a database of London school admission and discharge records as a proxy

[20] The estimates for the effect on occupational wages are qualitatively similar to those reported in the main specifications.

for school enrollment.[21] The admission registers provide information on the student's name, date of admission, date of birth, parents' names, and the name of the school. Although we can only infer attendance from enrollment, the admission registers provide a valuable (and perhaps unique) source of individual-level education data for 19th century England since the census did not include questions about education.

6 Conclusions

We have presented preliminary and tentative results suggesting that childhood illness has a long-run effect on socioeconomic outcomes. Oue main result suggests that hospitalized individuals were 5 to 10 percent less likely to be married in 1911 than their siblings. We also find some evidence that childhood illness was associated with lower occupational wages and physical skill requirements, but these findings are not robust across specifications. Poor childhood health may have restricted social mobility and contributed to income inequality in early-20th century England. While these results are not applicable to a developed-world context, they may also be relevant to large cities in the developing world today.

Our current results are limited by the lack of reliable metrics to rank occupational outcomes in 1911. We plan to supplement our analysis of occupational outcomes with two additional sources. First, we will assign History of Work Information codes (HISCO) to the occupational titles in the census and rank the occupational titles according to the HISCLASS socioeconomic status groups. Second, we will transcribe wages for approximately 100 occupation titles reported for major cities in England by the Board of Trade and the Labour Department around 1911. We will also extend our sample of childhood illness records by incorporating over 40,000 inpatient admissions to St. Bartholomew's Hospital in London between 1874 and 1906 for children aged 0 to 18 when admitted.

[21] Digitized admission and discharge records are publicly available on a genealogy website for 843 schools in one of the twelve boroughs of Inner London (including the City of London). The database contains over 1.7 million admissions between 1840 and 1911.

References

Abel-Smith, B. (1964): *The hospitals, 1800-1948: A study in social administration in England and Wales*, Cambridge, MA: Harvard University Press.

Abramitzky, R., L. P. Boustan, and K. Eriksson (2012): "Europe's Tired, Poor, Huddled Masses: Self-Selection and Economic Outcomes in the Age of Mass Migration," *American Economic Review*, 102, 1832–56.

Almond, D. (2006): "Is the 1918 Influenza Pandemic Over? Long-Term Effects of In Utero Influenza Exposure in the Post-1940 U.S. Population," *Journal of Political Economy*, 114, 672–712.

Almond, D. and J. Currie (2011): "Human Capital Development before Age Five," *Handbook of Labor Economics*, Volume 4, Part B, 1315–1486.

Barker, D. J. P. (1995): "The Fetal Origins of Coronary Heart Disease," *BMJ*, 311, 171–4.

Black, S. E., P. J. Devereux, and K. G. Salvanes (2007b): "From the Cradle to the Labor Market? The Effect of Birth Weight on Adult Outcomes," *The Quarterly Journal of Economics*, 122, 409–439.

Bleakley, H. (2007): "Disease and Development: Evidence from Hookworm Eradication in the American South," *The Quarterly Journal of Economics*, 122, 73–117.

——— (2010): "Malaria Eradication in the Americas: A Retrospective Analysis of Childhood Exposure," *American Economic Journal: Applied Economics*, 2, 1–45.

Case, A., A. Fertig, and C. Paxson (2005): "The lasting impact of childhood health and circumstance," *Journal of Health Economics*, 24, 365–389.

Currie, J. and E. Moretti (2005): "Biology as Destiny? Short and Long-Run Determinants of Intergenerational Transmission of Birth Weight," *National Bureau of Economic Research Working Paper Series*, No. 11567.

Currie, J., M. Stabile, P. Manivong, and L. L. Roos (2010): "Child Health and Young Adult Outcomes," *Journal of Human Resources*, 45, 517–548.

England, P. and B. Kilbourne (1988): *Occupational measures from the Dictionary of Occupational Titles for 1980 Census detailed occupations*, Inter-university Consortium for Political and Social Research.

Figlio, D. N., J. Guryan, K. Karbownik, and J. Roth (2013): "The Effects of Poor Neonatal Health on Children's Cognitive Development," *National Bureau of Economic Research Working Paper Series*, No. 18846.

Horrell, S., J. Humphries, and H.-J. Voth (2001): "Destined for Deprivation: Human Capital Formation and Intergenerational Poverty in Nineteenth-Century England," *Explorations in Economic History*, 38, 339–365.

Kingston University (2010): *HHARP: The Historic Hospital Admission Records Project*, Kingston upon Thames, Surrey: Kingston University, http://www.hharp.org.

Lam, L. L., E. Emberly, H. B. Fraser, S. M. Neumann, E. Chen, G. E. Miller, and M. S. Kobor (2012): "Factors underlying variable DNA methylation in a human community cohort," *Proceedings of the National Academy of Sciences*, 109, 17253–17260.

Lindahl, M., M. Palme, S. Sandgren Massih, and A. Sjogren (2012): "The Intergenerational Persistence of Human Capital: An Empirical Analysis of Four Generations," *IZA Discussion Paper No. 6463*.

Lomax, E. M. R. (1996): *Small and special: The development of hospitals for children in Victorian Britain*, London: Wellcome Institute for the History of Medicine.

Long, J. and J. Ferrie (2013): "Intergenerational Occupational Mobility in Britain and the U.S. Since 1850," *American Economic Review*, forthcoming.

Mill, R. and L. C. Stein (2012): "Race, Skin Color, and Economic Outcomes in Early Twentieth-Century America," *Working Paper*.

Miller, G. E., E. Chen, A. K. Fok, H. Walker, A. Lim, E. F. Nicholls, S. Cole, and M. S. Kobor (2009): "Low early-life social class leaves a biological residue manifested by decreased glucocorticoid and increased proinflammatory signaling," *Proceedings of the National Academy of Sciences*, 106, 14716–14721.

Minnesota Population Center (2008): *North Atlantic Population Project: Complete Count Microdata Version 2.0 [Machine-readable database]*, Minneapolis: Minnesota Population Center.

Nilsson, J. P. (2009): "The long-term effects of early childhood lead exposure: evidence from the phase-out of leaded gasoline," *Working Paper*.

Oreopoulos, P., M. Stabile, R. Walld, and L. L. Roos (2008): "Short-, Medium-, and Long-Term Consequences of Poor Infant Health," *Journal of Human Resources*, 43, 88–138.

Parman, J. (2010): "Childhood Health and Human Capital: New Evidence from Genetic Brothers in Arms," *Working Paper*.

——— (2012): "Childhood Health and Sibling Outcomes: The Shared Burden of the 1918 Infuenza Pandemic," *Working Paper*.

Rosenzweig, M. R. and J. Zhang (2009): "Do Population Control Policies Induce More Human Capital Investment? Twins, Birth Weight and China's "One-Child" Policy," *The Review of Economic Studies*, 76, 1149–1174.

Royer, H. (2009): "Separated at Girth: US Twin Estimates of the Effects of Birth Weight," *American Economic Journal: Applied Economics*, 1, 49–85.

Schürer, K. and M. Woollard (2003): *National Sample from the 1881 Census of Great Britain [computer file]*, Colchester, Essex: History Data Service, UK Data Archive [distributor].

Smith, J. P. (2009): "The Impact of Childhood Health on Adult Labor Market Outcomes," *Review of Economics and Statistics*, 91, 478–489.

Szreter, S. (2005): *Health and Wealth*, Rochester, NY: University of Rochester Press.

Tanner, A. (2007): "Choice and the children's hospital: Great Ormond Street Hospital patients and their families, 1855-1900," in *Medicine, Charity and Mutual Aid: The Consumptinoof Health in Britain, 1550-1950*, ed. by O. Abington, Great Britain: Ashgate Publishing Group, 135–161.

Waddington, K. (2000): *Charity and the London hospitals, 1850-1898*, New York: Royal Historical Society : Boydell Press.

Wohl, A. S. (1983): *Endangered lives: Public health in Victorian Britain*, Cambridge, Mass.: Harvard University Press.

Wolpaw Reyes, J. (2012): "Lead Policy and Academic Performance: Insights from Massachusetts," *National Bureau of Economic Research Working Paper Series*, No. 18327.

7 Appendix - Data Construction

7.1 Unique Patient Identifiers

Unlike most modern hospital patient records used by researchers, the HHARP admission records do not contain unique patient identifiers. Patients who were transferred from Great Ormond Street Hospital (GOSH) to the convalescent home Cromwell House (CH) can be tracked between record books based on dates of admission and discharge. The creators of the HHARP database identified these links and we take them as given. We manually assigned individual admission records to unique "patients" by requiring groups of records to match on the following criteria:

1. A "rough" match of first or middle name and surname. We allow for differences in the spelling of names, as well as transpositions of the first and middle names, or given and last names.

2. A 3-year band around the age at admission (+/- 1 year).
3. The records must also have either the same (i) street address of residence or (ii) cause of admission.

After matching the patient records to the census, we obtain some cases where multiple "patients" are linked to a single individual in the census. We show whether our results are robust to three methods of dealing with these observations: (1) consolidating these admissions into a single "patient", (2) treating each "patient" as a separate individual, and (3) discarding the observations.

7.2 Residential Address of Patients

In many cases, the HHARP records only include a street address of residence, while in others a geographic identifier such as a parish or registration district is also listed. We take as given the assignment of a corresponding registration district by the database creators. In the future, we plan to verify the district assignments and create a mapping from street addresses to registration districts in London from the street indexes to the 1881 and 1891 censuses. Street names often crossed multiple registration districts. We plan to identify the cutoff points based on the minimum and maximum values of street address numbers in each district for a given street name.

7.3 Census Search Criteria

Given that we consider the "patient" as the unit of observation for a match to the census, we search for all possible combinations of first names and surnames recorded across the set of admissions for a given patient. There is a concern that this method will bias the sample in favor of matching patients with multiple admissions. In future, we will also consider matching at the admission level.

The hospital records report the age at admission in years and months. The majority of observations for age in months are missing or heaped at 6 months, and thus we rely only on age in years. We assume a uniform probability of birth over the calendar year. We impute the year of birth as the midpoint of the one-year interval ending at the admission date minus the age at admission in years.[22]

If all admissions for a patient occurred before or after a particular census date, we use the county reported for the admission closest to that census date. Alternatively, if a patient was admitted before and after a census date, we search for the individual in the counties reported for the closest admissions on both sides of the census date.

7.4 Specification of Wildcard Names

The genealogy website allows users to enter wildcard characters in combination with full names ("?" to represent a single characters and "*" for multiple characters). For example, to search for "Joseph Addelsen" and "Joseph Addleton" we perform a single search that inputs "J?S?P?, Joseph" for the first name and "ADD*, Addelsen, Addleton" for the last name. Each search produces results that are consistent with the wildcard codes, as well as results similar to the actual name.[23] We group names with the same wildcard codes into a single search. The genealogy website requires at least three letters to be included if wildcards are used. With the exception of the first character in a string, and names with fewer than three consonants, we replace all vowels and any letters after the third consonant with wildcard characters.[24]

7.5 Census Matching Criteria

Around 8 percent of patients, or 25 percent of admissions, report ages at admission which correspond to multiple imputed years of birth. Many of the cases can be accounted for by inconsistencies between GOSH and CH records. We make the following assumptions regarding the imputed birth years:

[22] We use the same rule to impute the year of birth in the census.

[23] For example, names with phonetic similarity, equivalent soundex codes, and similar meanings or spellings.

[24] We drop any non-alphabetic characters such as the apostrophe in "O'Brien."

1. If a year of birth is imputed for at least 75 percent of admissions, we consider it to be the true birth year.
2. If this condition fails, we consider all imputed years of birth as equally plausible, and use the minimum distance between the census birth year and the set of birth years for the patient in the HHARP records.

7.6 Resolving Matches to Multiple Censuses

We identified the childhood household of patients by matching each patient to each of the 1881, 1891 and 1901 censuses. We created two samples based on separate criteria for selecting patients who were matches to multiple censuses.

1. Merge all matched records into a single file and choose the best match for each patient. If there are multiple matches of equivalent quality, choose the record in the census that was the closest to one of the patient's admission dates.
2. For each patient, select the matched record from the census that is the closest to one of the patient's admission dates.

A concern with the first method is that we may include a poorer quality match in our sample at the expense of a better quality match. Alternatively, the second method may be biased towards matching individuals who do not migrate. In the future, we plan to make use of other information in the census to determine whether a patient is matched to the same individual in different censuses.

7.7 Definition of Siblings

Although the majority of siblings are identified based on the presence of a parent in the household, we also attempt to identify siblings in cases where no parents are present in the household. We consider three categories of siblings:

1. Any set of individuals who report the same mother's name and father's name in the census are coded as siblings. We remain agnostic about whether individuals are full-

or half-siblings in cases where the father or mother is not present in the household. In cases where a group of individuals share the same mother, but only some report a father's name, we flag cases with a missing father as half-siblings.

2. Any set of individuals in a household who share a familial relationship to the household head (e.g. cousin, grandchild, nephew/niece, sibling).

3. Any set of individuals in a household who share a surname and a non-familial relationship to the household head (e.g. boarder, inmate, lodger). We restrict attention to individuals under the age of 16 since we are otherwise unable to distinguish between parents and children in these cases.

7.8 Definition of Match Quality Restrictions

In our regression specification we present four separate samples with varying degrees of restrictions on the quality of matches between patients and a childhood census, as well as between the childhood censuses and the 1911 census:

1. All matches: any observations which meet the minimum match criteria described above.

2. Precise matches: +/- one year of birth and no more than one string deviation in matching names.

3. Unique matches: no other results within two match distance units of the best match.

4. Precise and unique matches: the intersection of (2) and (3).

8 Tables

Table 1: Mean characteristics of households with children aged [0-11] and mother present

Characteristic	(1) England 1881	(2) London 1881	(3) Matched 1881	(4) Matched Pooled	(5) London 1881	(6) Matched 1881	(7) Matched Pooled
		All households				>=2 Male Siblings	
Panel A: Household characteristics							
Frequency of name combination	740.2	431.2	461.4	697.6	445.3	472.1	678.5
	(2043.8)	(1380.3)	(1184.3)	(1777.2)	(1273.6)	(1150.0)	(1657.5)
Total siblings	3.375	3.164	3.760	3.898	4.460	4.359	4.502
	(1.925)	(1.869)	(1.877)	(1.946)	(1.709)	(1.722)	(1.791)
Total male siblings	1.726	1.570	2.401	2.457	2.761	2.922	2.971
	(1.366)	(1.304)	(1.258)	(1.297)	(0.999)	(1.081)	(1.123)
Mean age of siblings	6.644	6.408	5.779	6.891	8.138	6.428	7.209
	(4.455)	(4.482)	(3.638)	(5.046)	(4.184)	(3.450)	(4.806)
Migrant	0.229	0.258	0.237		0.282	0.259	
	(0.420)	(0.437)	(0.425)		(0.450)	(0.438)	
Both parents present	0.942	0.936	0.941	0.897	0.940	0.948	0.908
	(0.234)	(0.244)	(0.236)	(0.303)	(0.238)	(0.222)	(0.286)
Mother's age at first birth	25.14	25.06	24.18		24.39	23.84	
	(13.97)	(14.31)	(4.850)		(13.97)	(4.551)	
Mother in Labor Force	0.307	0.241	0.226		0.222	0.210	
	(0.461)	(0.428)	(0.418)		(0.416)	(0.408)	
Panel B: Distribution of father's occupational categories							
Professional	0.0313	0.0402	0.0271		0.0400	0.0260	
	(0.174)	(0.196)	(0.162)		(0.196)	(0.159)	
Clerical	0.114	0.191	0.172		0.186	0.174	
	(0.318)	(0.393)	(0.377)		(0.389)	(0.379)	
Skilled	0.299	0.335	0.371		0.354	0.389	
	(0.458)	(0.472)	(0.483)		(0.478)	(0.488)	
Low-skilled	0.226	0.171	0.180		0.173	0.179	
	(0.418)	(0.376)	(0.384)		(0.378)	(0.383)	
Unskilled	0.212	0.139	0.135		0.140	0.129	
	(0.408)	(0.346)	(0.342)		(0.347)	(0.335)	
N	2769517	410012	2696	5458	181695	1965	4033

Notes: Columns (1) to (3), (5) and (6) present means for samples taken from the complete transcription of the 1881 census of England. Columns (1) to (4) include households with children under the age of 12, while (5) to (7) are restricted to households with at least two male siblings under the age of 12. Column (1) presents means for England; Columns (2) and (5) present means for London; Columns (3) and (6) present means for households in London with a child matched to the hospital records. Columns (4) and (7) pool matches to the 1881, 1891, and 1901 censuses. Columns (1), (2), and (5) exclude households with children matched to the hospital records. A migrant is defined as an individual residing in a county, at the time of enumeration, different from the county of birth. Panel B omits cases where an occupation is not reported or illegible, and the father is unemployed or not present. Missing cells are variables that have not been transcribed in the 1891 or 1901 censuses.

Table 2: Number of observations for patients and siblings in matched samples

		Number of Observations in 1911 Census by Match Quality				
	(1)	(2)	(3)	(4)	(5)	(6)
	Observations in Childhood Sample	All	All	Precise	Unique	Precise & Unique
Patients	5,430	1317	963	775	486	381
Siblings	10,627	1996	1,191	975	609	488
# Households with both	5,349	1308	955	769	484	380
N	16,057	3,313	2,154	1,750	1,095	869

Notes: Column (1) includes all male patients matched between the hospital records and either the 1881, 1891 or 1901 censuses, in households with at least one healthy male siblings. Column (2) includes individuals from households with at least one patient and one healthy sibling matched to the 1911 census. Columns (3) to (6) restrict the sample to patients and healthy siblings within 5 years of age of a patient in the household.

Table 3: Descriptive Statistics

Variable	(1) All	(2) All	(3) Precise	(4) Unique	(5) Precise-unique
Panel A: Characteristics from 1911 Census					
Hospital patient	0.399	0.445	0.440	0.441	0.435
	(0.490)	(0.497)	(0.496)	(0.497)	(0.496)
Age in 1911	29.91	28.93	28.92	27.97	27.88
	(7.236)	(6.149)	(6.141)	(6.093)	(6.070)
Birth order	3.076	3.130	3.118	3.098	3.080
	(1.772)	(1.692)	(1.697)	(1.615)	(1.600)
Married	0.556	0.539	0.526	0.496	0.465
	(0.497)	(0.499)	(0.499)	(0.500)	(0.499)
Head of household	0.544	0.526	0.514	0.480	0.450
	(0.498)	(0.499)	(0.500)	(0.500)	(0.498)
Panel B: Occupational skills and wages					
Eye-hand-foot coordination	4.414	4.423	4.412	4.446	4.437
	(0.568)	(0.571)	(0.578)	(0.561)	(0.570)
Finger dexterity	3.459	3.456	3.453	3.453	3.450
	(0.442)	(0.440)	(0.444)	(0.437)	(0.444)
Manual dexterity	3.189	3.193	3.187	3.202	3.201
	(0.372)	(0.377)	(0.377)	(0.374)	(0.375)
Motor coordination	3.367	3.362	3.358	3.365	3.360
	(0.353)	(0.354)	(0.355)	(0.358)	(0.362)
Numerical aptitude	3.364	3.354	3.360	3.363	3.366
	(0.565)	(0.555)	(0.549)	(0.543)	(0.542)
Physical demands scale	2.016	2.004	2.018	1.967	1.974
	(0.846)	(0.858)	(0.852)	(0.832)	(0.831)
Strength	2.574	2.558	2.570	2.544	2.543
	(0.746)	(0.747)	(0.749)	(0.728)	(0.732)
Ln (wage)	1.957	1.955	1.948	1.945	1.939
	(0.253)	(0.249)	(0.248)	(0.251)	(0.254)
Observations	2985	1950	1588	1001	797

Notes: Panel A includes variables reported in the 1911 census of England, as well as an indicator for the share of hospital patients in our sample. Panel B includes average occupational skill characteristics based on assigning each occupational string from the 1911 census to an occupational title in the 1980 US Census Dictionary of Occupational Titles (England and Kilbourne, 1988). Column (1) includes individuals from households with at least one patient and one healthy sibling matched to the 1911 census. Columns (2) to (5) restrict the sample to patients and healthy siblings within 5 years of age of a patient in the household. See Appendix section 7.8 for definitions of precise, unique, and precise-unique samples.

Table 4: Effect of childhood illness on Ln (wage)

VARIABLES	(1) All	(2) All	(3) Precise	(4) Unique	(5) Precise-unique
Panel A: specifications with indicator for hospitalization					
Hospital patient	-0.020	-0.008	-0.011	0.005	0.007
	(0.013)	(0.016)	(0.018)	(0.021)	(0.024)
Panel B: specifications with indicators for hospitalization by age group					
Patient (age 0 to 3)	-0.050**	-0.044	-0.048	-0.014	-0.002
	(0.023)	(0.027)	(0.031)	(0.035)	(0.043)
Patient (age 4 to 8)	-0.007	0.007	-0.010	0.013	-0.000
	(0.018)	(0.022)	(0.024)	(0.031)	(0.034)
Patient (age 9 to 13)	0.005	0.013	0.030	0.002	0.008
	(0.026)	(0.032)	(0.034)	(0.042)	(0.043)
N	2,996	1,958	1,592	1,005	799

Notes: Panel A and Panel B report coefficients from separate regressions. Each regression includes age in 1911, age in 1911 squared, and birth order (as proxied by age rank within the childhood census), and a sibling fixed effect. Standard errors are clustered at the sibling level.

Table 5: Effect of childhood illness on occupational skills

Dependent Variable:	(1) Finger dexterity	(2) Numerical aptitude	(3) Eye, hand, foot coordination	(4) Physical demand	(5) Manual dexterity	(6) Motor coordination	(7) Strength
Panel A: specifications with indicator for hospitalization							
Hospital patient	0.001 (0.039)	-0.044 (0.047)	0.085* (0.044)	-0.110 (0.069)	0.038 (0.033)	0.020 (0.033)	-0.100* (0.060)
Panel B: specifications with indicators for hospitalization by age group							
Patient (age 0 to 3)	-0.054 (0.059)	-0.053 (0.076)	0.032 (0.073)	-0.063 (0.116)	0.013 (0.053)	-0.048 (0.051)	-0.103 (0.100)
Patient (age 4 to 8)	0.016 (0.061)	-0.039 (0.064)	0.098 (0.063)	-0.069 (0.098)	0.037 (0.049)	0.045 (0.051)	-0.089 (0.087)
Patient (age 9 to 13)	0.052 (0.080)	-0.019 (0.110)	0.126 (0.101)	-0.256* (0.148)	0.063 (0.067)	0.056 (0.067)	-0.105 (0.124)
N				1,001			

Notes: Panel A and Panel B report coefficients from separate regressions. Each regression includes age in 1911, age in 1911 squared, and birth order (as proxied by age rank within the childhood census), and a sibling fixed effect. Standard errors are clustered at the sibling level. All columns report specifications using the "unique" from the previous tables.

Table 6: Effect of childhood illness on marriage probability

VARIABLES	(1) All	(2) All	(3) Precise	(4) Unique	(5) Precise-unique
	Panel A: specifications with indicator for hospitalization				
Hospital patient	-0.037*	-0.036	-0.038	-0.055*	-0.058
	(0.019)	(0.024)	(0.026)	(0.033)	(0.037)
	Panel B: specifications with indicators for hospitalization by age group				
Patient (age 0 to 3)	-0.030	-0.022	-0.025	-0.039	-0.026
	(0.033)	(0.041)	(0.047)	(0.055)	(0.066)
Patient (age 4 to 8)	-0.041	-0.046	-0.052	-0.077*	-0.095**
	(0.027)	(0.033)	(0.036)	(0.044)	(0.048)
Patient (age 9 to 13)	-0.010	-0.020	-0.019	-0.034	-0.032
	(0.040)	(0.051)	(0.056)	(0.070)	(0.080)
N	3,313	2,154	1,750	1,095	869

Notes: Panel A and Panel B report coefficients from separate regressions. Each regression includes age in 1911, age in 1911 squared, and birth order (as proxied by age rank within the childhood census), and a sibling fixed effect. Standard errors are clustered at the sibling level.

Table 7: Effect of childhood illness on household head probability

VARIABLES	(1) All	(2) All	(3) Precise	(4) Unique	(5) Precise-unique
	Panel A: specifications with indicator for hospitalization				
Hospital patient	-0.020	-0.009	-0.020	-0.024	-0.037
	(0.020)	(0.024)	(0.027)	(0.033)	(0.038)
	Panel B: specifications with indicators for hospitalization by age group				
Patient (age 0 to 3)	-0.028	-0.010	-0.025	-0.019	-0.025
	(0.032)	(0.040)	(0.046)	(0.054)	(0.064)
Patient (age 4 to 8)	-0.025	-0.024	-0.029	-0.075	-0.092*
	(0.028)	(0.034)	(0.037)	(0.048)	(0.054)
Patient (age 9 to 13)	0.024	0.023	0.014	0.054	0.053
	(0.040)	(0.051)	(0.055)	(0.067)	(0.076)
N	3,313	2,154	1,750	1,095	869

Notes: All columns report specifications using the “unique” from the previous tables. Standard errors are clustered at the sibling level.

Table 8: Effect of childhood illness on number of children

VARIABLES	(1) All	(2) All	(3) Precise	(4) Unique	(5) Precise-unique
	Panel A: specifications with indicator for hospitalization				
Hospital patient	-0.001	0.036	0.013	-0.012	-0.011
	(0.059)	(0.070)	(0.075)	(0.087)	(0.098)
	Panel B: specifications with indicators for hospitalization by age group				
Patient (age 0 to 3)	-0.068	-0.076	-0.063	-0.089	-0.089
	(0.101)	(0.112)	(0.123)	(0.138)	(0.155)
Patient (age 4 to 8)	0.027	0.038	0.029	-0.031	-0.005
	(0.080)	(0.098)	(0.106)	(0.133)	(0.141)
Patient (age 9 to 13)	0.032	0.171	0.067	0.169	0.140
	(0.121)	(0.147)	(0.151)	(0.166)	(0.189)
N	3,313	2,154	1,750	1,095	869

Notes: Panel A and Panel B report coefficients from separate regressions. Each regression includes age in 1911, age in 1911 squared, and birth order (as proxied by age rank within the childhood census), and a sibling fixed effect. Standard errors are clustered at the sibling level.

Table 9: Effect of childhood illness on probability of marriage and wages by age difference between patients and healthy siblings.

	(1)	(2)	(3)	(4)	(5)	(6)
	Difference in ages between patients and healthy siblings (years)					
VARIABLES	[0,2]	[0,5]	[0,8]	[0,2]	[0,5]	[0,8]
Panel A: Marriage						
Hospital patient	-0.118**	-0.055*	-0.040			
	(0.046)	(0.033)	(0.029)			
Patient (age 0 to 3)				-0.099	-0.039	-0.035
				(0.082)	(0.055)	(0.049)
Patient (age 4 to 8)				-0.126**	-0.077*	-0.051
				(0.063)	(0.044)	(0.040)
Patient (age 9 to 13)				-0.061	-0.034	-0.017
				(0.101)	(0.070)	(0.056)
N	531	1,095	1,402	531	1,095	1,402
Panel B: Ln(wage)						
Hospital patient	0.009	0.005	-0.015			
	(0.029)	(0.021)	(0.019)			
Patient (age 0 to 3)				0.017	-0.014	-0.037
				(0.049)	(0.035)	(0.033)
Patient (age 4 to 8)				0.012	0.013	-0.002
				(0.045)	(0.031)	(0.028)
Patient (age 9 to 13)				-0.034	0.002	-0.018
				(0.050)	(0.042)	(0.038)
N	483	1,005	1,282	483	1,005	1,282

Notes: All columns report specifications using the sample of "unique" matches from the previous tables (Column (4)). Standard errors are clustered at the sibling level.