NBER WORKING PAPER SERIES

EDUCATIONAL, LABOR-MARKET AND INTERGENERATIONAL CONSEQUENCES OF POOR CHILDHOOD HEALTH

Krzysztof Karbownik
Anthony Wray

Working Paper 26368
http://www.nber.org/papers/w26368

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
October 2019

We are grateful to Dr. Sue Hawkins for providing access to data from the Historical Hospital Admission Records Project (HHARP), and to Hardish Bindra at Paradigm Data Services for coordinating the transcription of the admission records from St. Bartholomew's and Guy's Hospitals. Wray is indebted to his dissertation committee members Joel Mokyr (chair), Joseph Ferrie, and David Dranove for encouragement and guidance. We also thank David N. Figlio, Guillermo Marshall, Werner Troesken, seminar participants at Emory University, the Federal Reserve Bank of Chicago, Hitotsubashi University, the Max Planck Institute for Demographic Research, the National Graduate Institute for Policy Studies (GRIPS), Northwestern University, and the University of Southern Denmark, conference participants at the World Congress of Cliometrics, the European Historical Economics Society, the Economic History Association, the Social Science History Association, the H2D2 Research Day at the University of Michigan, the NBER Children's Meeting, and the Public Health and Development Workshop at the University of Gothenburg for helpful comments and suggestions. Wray appreciates financial support from the Japan Society for the Promotion of Science KAKENHI Young Scientists B Grant Number J160100115 (PI: Wray), the Hitotsubashi Institute for Advanced Study (HIAS), Hitotsubashi University, the Northwestern University Economics Department's Eisner Fund and Center for Economic History, and the Economic History Association's Sokoloff Fellowship. The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

Educational, Labor-market and Intergenerational Consequences of Poor Childhood Health
Krzysztof Karbownik and Anthony Wray
NBER Working Paper No. 26368
October 2019
JEL No. I14,J62,N33

ABSTRACT

We study whether childhood health capital affects school attendance, long-run occupational outcomes, and intergenerational mobility. We address this question in the context of London, England during the late-nineteenth century using the inpatient admission records of three large hospitals linked to population census records, from which we identify household characteristics and the patients' siblings. Sibling fixed effects estimates indicate that boys with health deficiencies were 14.9 percent less likely to work in white collar occupations as adults and 13.9 percent more likely to experience downward occupational mobility relative to their fathers, in comparison to their brothers. This negative effect offsets 16.2 percent of the benefit of having a father in a high status occupation. We also explore medium-run mechanisms for both boys and girls, and find that poor childhood health reduced the likelihood of attending school by 2.5 and 4.1 percent, respectively.

Krzysztof Karbownik
Department of Economics
Emory University
Rich Memorial Building
1602 Fishburne Drive
Atlanta, GA 30322
and NBER
krzysztof.karbownik@emory.edu

Anthony Wray
Department of Business and Economics
University of Southern Denmark
Campusvej 55
Odense M - DK-5230
Denmark
wray@sam.sdu.dk

1 Introduction

Childhood health is an important determinant of future economic success (Currie 2009) while parental socioeconomic status in turn affects children's health (Case et al. 2002), potentially creating a cycle linking health and inequality across generations. An extensive body of research has shown that in utero conditions have persistent effects on human capital accumulation and labor market outcomes (Almond and Currie 2011b). The extant literature has also been largely successful at estimating the socioeconomic consequences of health during the first few years after birth (Almond and Currie 2011a). However, less is known about the long-run effects of health throughout childhood (Almond et al. 2018). A related literature has sought to explain the mortality decline and the demographic transition at the end of the nineteenth century (Cutler et al. 2006), but we know relatively little about health during life, differences in health between socioeconomic groups, and how health affected social well-being during this period (Costa 2015). Notably, health deficiencies in childhood could potentially have impacted the ability to attend school or enter into an apprenticeship (Horrell et al. 2001).

We make some progress in addressing these gaps in the literature by documenting that individual health capital between ages 0 and 11 affects school enrollment, occupational success, and intergenerational mobility in the context of London, England during the late-nineteenth and early-twentieth centuries. Our empirical approach combines three key elements. First, we obtain an individual-level measure of health that captures influences throughout childhood rather than during infancy alone. Second, we link individual childhood health status not only to medium-run schooling outcomes, but also to long-run occupational and intergenerational outcomes. And third, we combine these two inputs with a sibling fixed effects identification strategy.

We construct a measure of childhood health deficiencies by collecting historical inpatient records for the universe of children aged 0 to 11 admitted to three large London-area hospitals between 1870 and 1902.[1] Our hospitalization-based indicator of childhood health has several advantages. First and foremost, it yields estimates that are confounded by the positive effects of inpatient care to a lesser degree than studies using modern data, given that medical treatments and mitigation strategies were less effective historically. Moreover, in the context of our study, hospital admissions occur independently of any disease-specific policy intervention or population-wide mortality shock, which represent the typical sources of identifying variation in the literature. This distinction is important to the extent that the variation in health status in our study abstracts from any potential partial or general equilibrium effects that can arise when studying specific policies or shocks. Additionally, our measure of hospitalization captures health

[1] Other papers using hospital inpatient data to study the consequences of poor health include Arthi and Schneider (2017) and Doyle et al. (2019) in historical settings, and Currie et al. (2010) and Schwandt (2018) in modern contexts.

during childhood alone, in contrast to adult height, another common proxy for childhood health, which reflects the net influence of health throughout childhood and adolescence.[2] Lastly, it avoids the issue of recall bias in studies that use self-reported measures of health (Smith 2009).

In documenting long-run effects, we combine the individual-level hospital records with longitudinal census samples that contain demographic characteristics from childhood and socioeconomic outcomes during adulthood. We construct these samples from newly available complete-count records for the censuses of England between 1881 and 1911 which are provided by the Integrated Census Microdata (I-CeM) project. In doing so, we apply the methods for linking historical U.S. census data developed by Ferrie (1996) and refined by Abramitzky et al. (2012), Feigenbaum (2016), Mill and Stein (2016), and Bailey et al. (2017). We modify these methods to suit the U.K. context in which birthplace is non-standardized and reported at multiple levels of time-varying geographic boundaries (Schürer and Day 2019). Finally, in the empirical analysis, we create a sample of hospital patients and their siblings that pools together multiple hospital-to-census and census-to-census linkages.

We implement a sibling fixed effects identification strategy that compares hospital patients to their siblings who lived in the same household during childhood but were not hospitalized, as far as we can observe in the surviving records. Our estimates control for environmental factors common to the childhood household, as well as any time-invariant unobservables that may be correlated with both health status and economic outcomes. A limitation of the sibling fixed effects approach is that it cannot separately account for any time-varying parental responses such as compensatory or reinforcing behaviors that are related to the treatment. To the extent that these resource reallocation mechanisms are triggered by the initial health shock that led to the hospitalization, they would be captured by our long-run reduced-form estimates. Additionally, if the siblings were admitted to hospitals with no surviving records or if they were otherwise unhealthy, then our estimates should be interpreted as lower-bound effects.

We find that compared to their brothers, hospitalized males were 4.1 percentage points less likely to work in white collar occupations as adults, which offsets 16.2 percent of the advantage of having a father in a white collar occupation. Male patients were also 3.9 percentage points more likely than their brothers to experience downward occupational mobility, which accounts for 14.6 percent of overall downward mobility in England at the end of the nineteenth century. These results are robust to a variety of specification checks including issues related to selective

[2] Case and Paxson (2008, 2010) and Parman (2015a) use height as a proxy for childhood health and find that it is positively associated with cognitive test scores and levels of educational attainment, respectively. On the other hand, Case et al. (2005) document that conditional on height, the number of chronic conditions suffered at ages 7 and 11 have significant associations with education, which suggests that height does not capture all aspects of childhood health.

mortality, sample selection choices, and changes in the matching algorithm.

We then consider potential mechanisms for the long-run effects and show that hospitalized boys were 2.5 percent less likely to attend school compared to their non-hospitalized brothers in the 5-year period after the hospital admission. As is typically the case with historical data, we are unable to link hospital records for girls to long-run occupational outcomes due to name changes by women at the time of marriage. As such, it is rare to find evidence on the impact of childhood health for girls in a historical context.[3] However, we can ask how poor childhood health affects school enrollment for hospitalized girls compared to their sisters, and find that the effect on schooling is modestly larger than for boys at 4.1 percent.

This paper is among the few studies to use sibling fixed effects to identify the consequences of poor childhood health (Smith 2009; Currie et al. 2010; Beach et al. 2018; Hoehn-Velasco 2019).[4] Our study is the first to show that deficient childhood health as indicated by hospitalization has consequences that extend to older ages beyond the increased welfare participation in young adulthood found by Currie et al. (2010). Our results suggest that the consequences of poor childhood health also include lower occupational status as adults and a decline in relative status between generations. Similar to our findings, Hoehn-Velasco (2019) estimates that exposure to a preventative health program has a positive impact on adult earnings. Our study is also one of the relatively few papers on the long-run effects of childhood health measured by the disease environment beyond the early-life period. In particular, Bütikofer and Salvanes (2019) explore the impact of a tuberculosis testing and vaccination campaign, and Gensowski et al. (2018) use data on childhood hospitalizations for polio to study the effects of physical disability from the quasi-random incidence of paralytic polio.[5] In contrast, our historical urban setting and the causes of admission reported in the hospital registers are characterized by variation in health status that stems from a broad range of conditions, which arguably increases the external validity of our estimates. Moreover, the consequences of the historical health environment remain relevant in modern times given that similar conditions are still prevalent in the developing world, especially in rapidly urbanizing cities (Fogel 2004; Floud et al. 2011; Currie and Vogl 2013).

Our paper also connects to a literature that uses the disease environment around the time of birth as a proxy for early-life health to evaluate its long-run consequences. Previous studies typ-

[3]This significant omission from our knowledge base is now being remedied by the ongoing Longitudinal, Intergenerational Family Electronic Micro-Database (LIFE-M) Project and other related projects making use of birth and marriage registers (Bailey et al. 2017).

[4]The use of sibling fixed effects among studies based on historical census data is also limited, but growing with the recent availability of complete count data (Abramitzky et al. 2012; Parman 2015a; Mill and Stein 2016).

[5]Other studies have evaluated the impact of policy changes during childhood such as expansions in childhood health insurance coverage (Goodman-Bacon 2017; Brown et al. 2019) and improvements in living conditions (Gould et al. 2011).

ically exploit cross-sectional variation in mortality from infectious diseases such as hookworm (Bleakley 2007), malaria (Barreca 2010; Venkataramani 2012; Hong 2013), pneumonia (Bhalotra and Venkataramani 2012), typhoid fever (Beach et al. 2016), or yellow fever (Saavedra 2017). Many studies obtain causal estimates by interacting this cross-sectional variation with quasi-random temporal changes in mortality due to the discovery of antibiotics (Bhalotra and Venkataramani 2012; Zhang 2014; Lazuka 2019) or public health campaigns that sought to eradicate tropical diseases (Bleakley 2007, 2010; Venkataramani 2012; Baird et al. 2016).[6] Our results on intergenerational mobility also provide context to prior literature that has estimated intergenerational elasticities and rates of occupational mobility for the historical period that we study (Long and Ferrie 2013; Long 2013; Clark 2014; Olivetti and Paserman 2015; Pérez 2019). In particular, our results indicate that poor childhood health can account for a non-trivial share of overall mobility by reducing upward mobility and increasing downward mobility.

2 Historical Background

Mid-nineteenth century England was characterized by a minimal degree of effective medical treatment (Lomax 1996) and limited knowledge of preventative health behaviors (Mokyr 2000; Worboys 2000). Child mortality due to infectious diseases such as scarlet fever, typhoid fever, cholera, tuberculosis, whopping cough, and smallpox declined from the 1860s onward (Mercer 2014), but the perception of sickness was defined by the daily discomforts of colds, headaches, and diarrhea (Hardy 2001). Factors such as overcrowded housing, inadequate sanitary conditions, resource constraints, and a lack of proper nutrition or medication arguably contributed to poor childhood health (McKeown 1976; Wohl 1983; Szreter 2005).

During the second half of the nineteenth century, medical care for children transitioned from informal home care to formal institutional settings. This shift was precipitated by changing attitudes towards the care of children, the growing professionalization of medical care, and the increasing availability of medical technology. Until then, the health of children was not viewed as a state responsibility in England. Even at general hospitals that received no state funding and financed their operating costs out of private endowments, the admission of infants was discouraged on the basis of the beliefs that the baby would suffer ill-effects from the separation from

[6] Other papers that evaluate long-run outcomes obtain exogenous variation in early-life health conditions from the roll-out of policy interventions. Many studies focus on interventions during infancy or very early in life such as access to infant health care centers and their impact on socioeconomic outcomes (Bütikofer et al. 2019) or the introduction of a home visiting program and its impact on health outcomes (Hjort et al. 2017). Other childhood health interventions shown to have affected socioeconomic outcomes include improvements in nutrition (Adhvaryu et al. 2019). Studies that examine health outcomes exploit variation in early-life health that stems from the introduction of sulfa drugs (Jayachandran et al. 2010) or the eradication of malaria (Hong 2007).

their mother and that hospitalization challenged parental authority. Furthermore, until the late-nineteenth century, few general hospitals had specialized wards or staff to admit child patients (Lomax 1996).

Children's hospitals were established to provide specialized medical care and surgical treatments for sick children. Doctors at the hospitals could develop knowledge of rare childhood diseases and train specialized nurses and medical students. The hospitals were founded with the aim of spreading middle-class values and providing education about preventative health behaviors to the "deserving" or "independent" poor (Hawkins and Tanner 2013). On the other hand, destitute paupers were denied admission to hospitals and sent to poor law infirmaries where they could not be refused (Brunton 2004). By the end of the nineteenth century, children's hospitals no longer faced the "stigma of charity," nor were they feared as places where patients would die of hospital-acquired infections. Hospitals came to be viewed as the most suitable venue for treating acutely and chronically ill children and those requiring surgery. Moreover, the willingness to seek admission to the hospital extended beyond the working poor to include the middle and even the privileged upper classes (Lomax 1996).

The first children's hospital in Britain, the Hospital for Sick Children at Great Ormond Street (GOSH), opened in 1852 in London and by 1870, London had seven children's hospitals. GOSH opened with a capacity of 10 beds and would grow to 62 beds by 1864, before expanding to 120 beds in 1877 and 240 beds in 1914 (Franklin 1964; Hawkins and Tanner 2013). Among children age 0 to 11, the number of inpatients at GOSH per 1,000 in the London population increased from 0.98 in 1881 to 1.90 in 1901, while inpatients per 1,000 at St. Bartholomew's Hospital (Barts), one of the largest general hospitals in London, declined from 1.05 to 0.67 during the same time period, as children's hospitals became the preferred location for treating children.

Admission to the hospital inpatient department was selective as medical attendants had considerable authority over the types of cases admitted and could prioritize the admission of outpatients with acute illness or rare childhood diseases. In the case of children, the parent accompanying a child to the hospital was first screened by a clerk to determine the family's ability to pay for the hospital stay, while the child was examined by the house surgeon or physician on duty to determine suitability for admission as an inpatient.[7] Parents of children who were not admitted received some medical advice (Hawkins and Tanner 2013). By regulation, hospitals were typically expected to exclude chronic or incurable cases, to avoid having a bed occupied for a lengthy period of time, and infectious diseases to limit the number of deaths at the hospital. However, as we observe in our data, many such cases were admitted in practice.

[7]During the 1880s, early in our study period, the large general hospitals began collecting admission fees from patients who could afford the payment, but still admitted those who could not (Higgs 2009).

Despite improved diagnostics following gains in knowledge about bacteriology and the increased prevalence of autopsies, pediatrics was largely limited to convalescent care. Medical patients benefited from bed rest, nursing care, and an adequate diet while admitted to the hospital. Especially among impoverished patients, children were exposed to a much more sanitary environment than the overcrowded conditions at home (Higgs 2009). From the 1880s onward, hospital practices advanced in the area of surgical procedures with improved knowledge of the bacteria that caused surgical infections and better training of nurses, which enabled the control of post-operative sepsis. Hospitals also limited the spread of infection by establishing isolation wards. As the mortality risk declined, hospital began performing a wider range of surgeries that could improve the quality of life for patients, including the removal of tubercular glands or the appendix, and the repair of congenital malformations such as cleft palate or club foot (Lomax 1996).

3 Data and Descriptive Statistics

Our paper draws on a new data set of individual-level hospitalization records that we compile from the late-nineteenth and early-twentieth century inpatient admission registers of three large hospitals in London, England. We link the hospital admission records to the complete-count Population Censuses of England to bring together the data on childhood health with information on family structure, socioeconomic status, demographics, and school attendance during childhood, as well as occupational and intergenerational outcomes during adulthood.

3.1 Inpatient hospital records

The first branch of our data set consists of inpatient hospital admission records, which provide a measure of health during childhood. We digitized and transcribed the inpatient admission registers from two of the four large general hospitals in London: St. Bartholomew's Hospital (Barts) and Guy's Hospital (Guy's). We supplement these data with the admission records from the Hospital for Sick Children at Great Ormond Street (GOSH), the largest and oldest children's hospital in London.[8] A map of central London in Figure A1 shows the locations of the hospitals in the data as well as other nearby general and children's hospitals, and highlights that admission records have survived for only four out of the ten largest general hospitals and two of the five largest

[8]The records from the Hospital for Sick Children at Great Ormond Street were transcribed by volunteers in London from inpatient admission registers as part of the Historical Hospital Admissions Research Project (HHARP; Kingston University 2010). We thank Dr. Sue Hawkins for sharing these data.

children's hospitals in London. We discuss how selection into hospital admission and selective survival of records may impact our results in Section 3.5.

The hospital records contain detailed information about the patient and the characteristics of the admission. Figure A2 presents a sample page of an individual admission register from Barts Hospital, which is similar to the records for the other hospitals in our sample. An individual entry includes the patient's full name, age in years, and residential address, in addition to a description of the patient's cause of admission, the dates of admission and discharge, the name of the attending physician or surgeon, and an indication of whether the patient died in the hospital. For all hospitals, we observe both male and female patients. We use the information contained in the hospital records to link individuals to the census and to construct a health deficiency index that measures the severity of the cause of admission.

Children were admitted to the hospital as inpatients for a wide-ranging set of conditions. Table A1 lists the most common causes of admissions to the hospitals among all hospital patients, all male patients, and those in the sample of male patients used in the main analysis.[9] Overall, the death of an inpatient while admitted to the hospital was not an uncommon occurrence among the cohorts in our study, with an average mortality rate of 11 percent, and a mortality rate as high as 26 percent for individuals admitted before the age of two (Figure A3).[10] Additionally, the average mortality rate was relatively constant over the sample period (Figure A4). As an alternative treatment variable to the hospitalization indicator, we construct a gender-specific childhood health deficiency index that is based on the unexplained portion of the inpatient mortality after removing the influence of the hospital, age at admission, and year of admission using a procedure described in appendix B. This variable is standardized on a 0 to 1 scale and has a mean value of 0.28 and standard deviation of 0.10 among the hospital patients in the estimation sample. We interpret causes of admission with higher residual mortality as having greater severity and thus the health deficiency index represents a proxy for childhood health capital that varies across causes of admission with higher values indicating more severe diagnoses.

The analysis restricts attention to the admission records of male and female patients from the 1870 to 1890 birth cohorts who were admitted between 1870 and 1902, and hospitalized between the ages of 0 and 11 years.[11] Although older children were also hospitalized at the general hospitals, we choose age 11 as the cutoff for inclusion in the sample since, as a rule, the children's

[9] Throughout the paper we present results separately by gender unless otherwise noted.

[10] While comparable estimates of inpatient mortality rates from the nineteenth century are limited, hospital records from the Royal Hospital for Sick Children in Glasgow, Scotland indicate that the inpatient mortality rate between 1890 and 1899 was lower at that hospital at 5.9 percent (Cullis and Young 2013).

[11] The number of observations is balanced across age groups and admission periods with the exception of admissions between 1870 and 1875 given that we only include individuals born from 1870 onward (Figure A5).

hospitals did not admit patients at older ages.[12] Furthermore, children age 12 and older were much less likely than younger individuals to be living with their parents (Figure A6), which is a crucial requirement for the sibling-fixed effects estimation strategy that will be described in Section 4.

3.2 Linked Census of England complete counts, 1881 to 1911

We estimate the effects of childhood health capital on school attendance, long-run labor market outcomes, and intergenerational mobility by linking the inpatient admission records to childhood household characteristics and socioeconomic outcome variables from the complete-count files of the Population Censuses of England provided by the Integrated Census Microdata (I-CeM) project.[13] First, we locate the hospital patients in their childhood households in the 1881 through 1901 censuses, and collect information on their father's occupational title, the identity of their siblings, the educational status of patients and siblings, and the places of birth (parish and county, or country for foreign births) of all family members. Then, we link individuals across censuses to observe patients and their siblings as adults in the 1901 and 1911 censuses, and use occupational titles to evaluate their long-run labor market outcomes. Since women changed their surnames at marriage, the analysis of long-run outcomes is restricted to men.[14] We also create a separate hospital-to-census linked sample of both males and females to explore the role of schooling as a mechanism, which we discuss in Section 3.4.

More specifically, for the long-run analysis, we match admission records to individuals enumerated in each of the 1881, 1891 and 1901 censuses, which potentially generates multiple linkages for each admission record. We use first and last name and the approximate birth year implied by the age at admission or age at enumeration as matching variables. In the absence of a unique patient identifier in the admission registers, we identify unique individuals based on shared char-

[12] The founders of the children's hospitals initially intended for the hospitals to treat patients aged 2 to 12, leaving mothers to care for sick infants and general hospitals to admit older children (Lomax 1996). In practice, an upper age limit of 10 to 12 years was enforced with some exceptions, but medical staff typically disregarded the ban on under-two admissions (Hawkins and Tanner 2013), which would account for over 30.1 percent of GOSH patients by 1900. In Section 6 we show that our main results are robust to excluding age 0 to 1 admissions from the sample. At GOSH, a few 12 to 16 year olds were admitted, but as they only accounted for 0.7 percent of admissions during the sample period (1870 to 1902), we exclude these individuals from the sample with no bearing on the main results.

[13] The digitized complete-count census records of Great Britain for 1851 to 1911 are available to download from the I-CeM project, but do not include individual names or street addresses. We are among the first researchers to obtain access to these restricted-use variables through a Special License from the UK Data Archive. The complete-count censuses from Britain represent the second largest historical census microdata collection after the U.S. (Ruggles 2014).

[14] In an earlier version (Karbownik and Wray 2015), we explored the possibility of conducting an additional linkage to marriage certificates to incorporate women in the long-run analysis. However, we found that the digitized collections of marriage certificates for London are incomplete and that adding a fourth linkage would have reduced our sample size drastically.

acteristics across admissions using a procedure described in appendix A.3. Separately, we link the universe of males from the cohorts of interest in each of the 1881 and 1891 censuses to both the 1901 and 1911 censuses, in addition to linking the 1901 to the 1911 census. In the census-to-census matching applications, we use place of birth as a linking variable, which allows us to match individuals who migrate within England between childhood and adulthood. In both cases, we use an iterative matching algorithm that draws on elements of Ferrie (1996), Abramitzky et al. (2012), Mill and Stein (2016), and Feigenbaum (2016).[15] Appendix A provides a detailed description of the hospital admission-to-census and census-to-census linkages, including how we consolidate multiple hospital-to-census matches and arrive at a final set of unique patient identifiers and sample individuals.

Table 1 presents baseline sample sizes and linkage rates for the three hospital admission-to-census linkages (to the 1881, 1891, and 1901 censuses) in columns 1 to 3. In each case we consider hospital admissions by male patients that occurred within ten years of census enumeration. Column 4 shows overall linkage rates for hospital admissions from pooling together the separate admission-to-census links, on which our analysis is based. In panel A we find a unique match to a census record for 34.3 percent of hospital admissions, fail to find any match for 3.5 percent of admissions, and exclude the remaining 62.2 percent of cases for which there are multiple potential matches. The unique match rate is similar to the match rates in other studies (Abramitzky et al. 2012; Long and Ferrie 2013). Further restricting the unique matches to households in which a male sibling is present only eliminates an additional 2.1 percent of all admissions, leaving us with 32.2 percent of the baseline sample. Panel B presents consolidated linkage rates for each childhood census year from combining matches to either the 1901 or 1911 censuses during adulthood. We match 21.9 percent of the baseline sample to census records in both childhood and adulthood, while in 10.9 percent of cases, we are also able to link a male sibling between censuses. After eliminating observations with missing cause of admission or occupational outcome, our final sample includes 2,320 hospital admissions or 8.2 percent of the baseline sample.

[15] While previous applications of the historical census linkage methods have often involved U.S. data, U.K. censuses are characterized by important differences in comparison to the U.S., which necessitate different approaches to data linkage. U.K. censuses report both county and parish of birth, whereas the U.S. only reports state of birth, which are larger geographic units by population than either parish or county. While state and county are comparable in terms of data quality and number of unique entries, birth parishes have boundaries that change over time and are not consistently reported across censuses. Thus, in the main analysis, we match on birth county when linking records across censuses. Countries other than the U.S. that have been studied using linked historical census data include the U.K. (Long 2013; Long and Ferrie 2013), Norway (Abramitzky et al. 2012), and Argentina (Pérez 2019).

3.3 Occupational outcomes

The occupational titles reported in the 1901 and 1911 censuses contain information on an individual's social class and provide the basis for a measure that compares the occupational attainment of hospital patients and their siblings as working adults. We also construct measures of intergenerational mobility that compare the occupational success of patients and siblings conditional on their father's occupational status when they were children, which we obtain from the census during childhood. The latter measure indicates the extent to which poor childhood health hinders the intergenerational transmission of status.

The I-CeM project has assigned a Historical International Standard Classification of Occupations (HISCO) code to each of the unique occupation strings in the complete-count census files (van Leeuwen et al. 2002; UK Data Archive 2014). We rank the socioeconomic status of the occupational titles according to the Historical International Social Class Scheme (HISCLASS), which maps each of the 16,000 HISCO occupation codes to one of twelve social classes (van Leeuwen and Maas 2011). The assignment of the HISCLASS category is based on the extent of supervision and skill level required by the occupation, whether the occupation is manual, and by the economic sector of the occupation.

In the main empirical analysis of individual occupational success, we consolidate the HISCLASS ranking into four classes, which we refer to as white collar (e.g. clerk; HISCLASS 1-5), skilled (e.g. cabinet maker; HISCLASS 6-8), semi-skilled (e.g. house painter; HISCLASS 9), and unskilled (e.g. general laborer; HISCLASS 10-12). Table A2 lists the most common occupations in each of the four groups that we use to construct binary dependent variables in the analysis of unconditional occupational attainment. When we turn to the intergenerational outcomes that compare the occupational status of patients and siblings to their fathers, we follow Abramitzky et al. (2011) and base the comparison on a consolidated HISCLASS ranking with seven social classes. In comparison to the ranking with four classes, the seven-class measure has three differences. First, among white collar workers, managers and professionals are separated from clerical and sales personnel. Second, farmers and fishermen are placed in a separate group ranked between skilled and semi-skilled workers.[16] Third, low-skilled and unskilled farm workers are ranked below other unskilled workers.

[16] Given that our sample is predominantly urban, we have very few farmers and unskilled farm workers from classes four and seven, respectively.

3.4 School attendance

We consider the role of human capital accumulation as a potential mechanism that accounts for the long-run effects on occupational status by constructing a measure of school attendance for children age 5 to 10. These individuals were subject to compulsory schooling under the Elementary Education Act of 1880.[17] In the absence of specific questions on education in the historical censuses of England, we use information contained in the occupational field in the 1881 and 1891 censuses. In these census years, enumerators were instructed to record "scholar" in the occupational field for children and young persons above the age of five who attended school on a daily basis. Furthermore, enumerators typically did not record an occupation for children younger than 12 years old, and thus, being regarded as a"scholar" arguably provides a reliable indicator of school attendance.

The difference in how enumerators recorded information in the occupational field for individuals age 12 and below is born out in the census data. Figure A7 plots the school enrollment rates and the labor force participation rates for children aged 5 to 18 in the 1881 census. The census-based measure suggests that compliance with compulsory schooling was relatively high as 64 to 82 percent of children aged 5 to 10 were recorded as a "scholar," while fewer than 1 percent of children age 10 and below reported a gainful occupation. The latter helps to rule out occupational health hazards as a potential mechanism for our findings. When constructing the linked sample, we match patients to either the 1881 or the 1891 censuses if they were admitted to the hospital prior to enumeration and they were ages 5 to 10 years old at the time of the census.[18] This implies that the estimation sample includes individuals from the 1871 to 1876 birth cohorts linked to the 1881 census and individuals from the 1881 to 1886 birth cohorts linked to the 1891 census.

3.5 Sample selection

In this section we characterize the sample selection into hospital admissions as well as in relation to data availability and methodological choices at different stages of the analysis. Although admission registers from the nineteenth century have survived to the present for only a handful of hospitals in London, those that do remain are among the largest hospitals in their respective categories and accounted for a sizable fraction of the market for hospital care at the time (Chatto and Windus 1897). Table A3 shows that Barts and Guy's Hospitals accounted for 25 percent of

[17] The Elementary Education (School Attendance) Act of 1893 raised the minimum school leaving age to 11, while an amendment to the Act in 1899 further raised the compulsory age to 12, but these changes occurred after the latest census year in which we observe school attendance, 1891, and thus we exclude 11 and 12 year olds from this analysis.

[18] The 1901 census did not report whether a child was a "scholar" and did not report years of education. Thus, we do not link hospital patients to the 1901 for this exercise.

inpatient admissions among the twelve largest general hospitals in London, while GOSH covered a large share of the children's hospital market with 29 percent of inpatients. The hospitals in our sample had similar market shares in terms of outpatients and hospital beds.

As we cannot speak to hospitals without surviving records, we first consider how registration districts in which the sample hospitals were located and neighboring districts differed from the rest of London. We define a hospital's catchment area as the set of registration districts in which the largest share of inpatients resided and which together accounted for at least 50 percent of admitted patients. The majority of patients admitted to the general hospitals, Barts and Guy's, resided in districts immediately surrounding the hospital, whereas a specialty children's hospital such as GOSH had a much larger catchment area and typically admitted patients from a wider part of London. Table A4 presents descriptive statistics from the 1891 Census of England for the catchment areas of each of the three hospitals in the sample, in comparison to the rest of London.[19] Since Guy's Hospital was located in the poorer borough of Southwark to the south of the Thames River, it is not surprising that the share of unskilled fathers in its catchment area (20 percent) was much larger that the shares for Barts and GOSH (11 and 12 percent, respectively), and the share for the rest of London (16 percent). Aside from the differences in average occupation status, the hospital catchment areas were similar to one another in terms of observable measures from the census, such as the share of children aged 0 to 4 or 5 to 11 in the population, sibship size, the share of children (age 0 to 11) living with their parents, the share of married households, or the share of immigrants.

Next, Table A5 examines how household socioeconomic status determined selection into hospital admission among children age 0 to 5 and residing in London at the time of census enumeration. We consider inpatient hospital admissions that occurred up to 10 years after census enumeration. We find evidence of an SES gradient as children with semi-skilled or unskilled fathers were more likely to be admitted to the hospital than children with white collar fathers, regardless of whether we look across London, within registration districts or within parishes.[20] The finding that children from white collar backgrounds were less likely to be admitted is consistent with the fact that upper classes could afford medical treatment and relied on general practitioners who operated private clinics and treated patients in the privacy of their homes (Carpenter 2010). Sample selection arising from differences in hospital catchment area characteristics or household SES relative to the population will not affect the internal validity of our estimates to the extent that it is absorbed by the sibling fixed effects.

[19]We use the 1891 census since it is closest to the midpoint of admission years in our sample. The comparisons of hospital catchment areas are similar when based on the 1881 or 1901 censuses.

[20]We find a similar SES gradient when the outcome variable is hospital admissions that occurred in the 10 years prior to census enumeration, but choose not to include these specifications since the father's occupational status will be endogenous to the child's health deficiency.

However, within-household selection due to parents' differential sending of their children to the hospital based on health status could affect internal validity. Columns 1 to 4 of Table A6 explore how sibling-specific factors affect the likelihood of hospitalization. We consider samples of census records linked forward to hospital admissions within 10 years of enumeration and assume that unlinked individuals were not hospitalized. Using a sibling fixed effects specification, we find that first-born males and females regardless of parity were less likely to be hospitalized. To the extent that parents valued first-born males greater than other children, our sample of hospitalized children will be negatively selected due to the first-born advantage present even in our historical data, and thus we include birth order fixed effects in our empirical specification to control for this potential bias. We abstract from gender-specific selection by restricting to a male-only sample in the main analysis. Despite evidence of parity-based selection into hospitalization it is reassuring that the selection appears to be unrelated to health, as columns 5 to 8 show that among hospital patients, first-born status and gender are unrelated to the health deficiency index at admission which is our proxy for severity.

Another potential source of sample selection bias is differential rates of linkage from the hospital records to the childhood censuses by health status. Table A7 examines how the health deficiency index of a patient admitted to the hospital affects the likelihood of finding a unique match in the census immediately following the admission. Columns 1 to 3 present results for linkages to each of the 1881, 1891, and 1901 censuses separately, while column 4 pools the samples and considers the impact on linkage to any census. We find evidence of positive selection into the sample as patients with a worse health deficiency index at admission are less likely to be linked to a census. The magnitude of these estimates is quite small with a 1 s.d. (0.10) increase in the health deficiency index reducing the likelihood of a match by about 0.6 percentage points or 2.2 percent. Furthermore, Figure A8 shows that the distribution of the health deficiency index in the universe of hospital admissions is similar to the distribution in the sample used in the analysis. It is primarily patients admitted for conditions with a very high health deficiency index (i.e. high in-hospital mortality) who are naturally less likely to be matched. This positive selection would downward bias our estimates, leading to a lower-bound interpretation, but as noted above the magnitude of the selection appears small.

Lastly, we consider sample selection due to differential linkage rates between patients and siblings. In panel B of Table A8 we compare match rates from censuses during childhood to censuses during adulthood for patients compared to their siblings. In general patients are more likely to be linked, though the results are only significant for matches from the 1891 census and the magnitudes are small at 2.4 pp (4.1 to 4.3 percent relative to the mean). Since patients in this analysis have already been matched from hospital records to a census during childhood, this is a plausible factor making patients more likely to be matched.

Collectively, the evidence on sample selection suggests that the patients in our linked empirical sample are positively selected relative to their brothers. Thus, our point estimates should provide a lower bound for the role of childhood health in long-run occupational and intergenerational success. Additionally, as noted above, many of the selection coefficients are small in magnitude even if statistically significant.

4 Empirical Specification

We use a sibling fixed effects model to estimate the effects of childhood health deficiencies on school attendance, occupational class, and intergenerational occupational mobility for male patients and their brothers. We estimate the following regression for child i from household j:

$$Y_{ij} = \beta HealthCapital_{ij} + \gamma X_i + \alpha_j + \varepsilon_{ij} \tag{1}$$

where Y_{ij} is a schooling, occupational, or intergenerational outcome for individual i from childhood household j, $HealthCapital_{ij}$ is a measure of deficiency in health capital, X_i is a vector of individual characteristics, α_j denotes a sibling fixed effect that captures unobservable time-invariant determinants of the outcomes specific to a childhood household, and ε_{ij} is a heteroskedasticity-robust error term clustered at the childhood household level that represents sibling-specific unobserved characteristics.

The variable $HealthCapital_{ij}$ is proxied by an indicator for hospitalization during childhood or the health deficiency index described in appendix B. The coefficient of interest is β, which can be interpreted as the differential effect of lower childhood health capital for a hospitalized child compared to his brother. The vector of individual characteristics X_i includes age-by-census year and birth-order fixed effects, an indicator for the older sibling in each pair of observations from a household, standardized measures of first name frequency and an interaction of first name and surname frequency, and indicators for match quality (exact matches on first name, surname, and birth parish between censuses, or cleaned birth places).[21]

The analysis of individual occupational success uses measures based on collapsing the HISCLASS ranking into four classes as described in Section 3.3. The outcome variables Y_{ij} represent one of three indicators: the probability of working in a white collar occupation, the probability of working in a white collar or skilled occupation, and the probability of working in a white collar, skilled, or semi-skilled occupation. In the last case, the lowest status category of unskilled

[21] We include age-by-census year fixed effects to account for the fact that a 25 year old observed in the 1901 census comes from a different birth cohort than a 25 year old observed in the 1911 census. We do not include surname frequency in levels since it is absorbed by the household fixed effects.

occupations constitutes the comparison group. These outcome variables can be thought of as a cumulative distribution function for occupational class.

When the focus turns to intergenerational mobility in occupational attainment, we study the effects on two additional indicator variables. The first measure indicates the probability of attaining a lower occupational class than one's father and the second variable identifies individuals who attain the same occupational class or higher than their fathers. Among patients in the main analysis sample 36 percent have the same status as their father, while for siblings this share is 37 percent. These intergenerational outcomes represent measures of occupational success relative to family endowments and we consider the stability of status across generations as a positive outcome.

Given that upward mobility is not possible if a father has the highest occupational status, and downward mobility is not feasible when a father is in the lowest class, we exclude these cases from the sample. Furthermore, the effects on a downgrade in occupational class are only identified from cases in which the patient attains a lower occupational class than his father, but the non-hospitalized sibling does not, and similarly for an upgrade in status. In our main sample, 74 percent of patient-sibling pairs have the same outcomes relative to their fathers and thus do not contribute to the identification. Reassuringly, an indicator for households with patient-sibling variation in outcomes is uncorrelated with potentially confounding explanatory variables (Miller et al. 2019).[22]

5 Main Results

We begin by analyzing the effects of childhood health deficiencies on long-run occupational status and intergenerational mobility using the main estimation sample of male patients and their brothers linked to census records during adulthood. We put the magnitudes of the estimates into context by estimating the intergenerational transmission of occupational status and computing overall rates of occupational mobility in the population as benchmarks. Then, we turn to an examination of potential mechanisms and highlight the role of human capital accumulation by considering school attendance as an intermediary outcome. We also conduct heterogeneity anal-

[22] An indicator for patient-sibling pairs with variation in outcomes relative to their fathers is uncorrelated with father's age categories, sibship size, characteristics of the patient's linkage to the census, and whether the household is located in the hospital's catchment area. Households with fathers in skilled or semi-skilled occupations are more likely to have variation in outcomes than those with fathers in unskilled occupations, which reflects the greater mobility in and out of the middle of the occupational distribution. Given our results reported in section 5, we naturally find that households with patients who had a worse health deficiency index are more likely to have variation in outcomes relative to the father. These patterns are similar for the individual occupational outcomes.

ysis to provide further insight into the channels linking poor childhood health to labor market outcomes in adulthood.

The main regression results for the long-run occupational and intergenerational outcomes, based on sibling fixed effects, are presented in Table 2. Panel A presents estimates in which the treatment of interest is an indicator variable for hospitalization, while panel B contains estimates in which the treatment is the health deficiency index that exploits variation in severity across causes of admission. Columns 1 to 3 display the results for the occupational outcomes, while columns 4 and 5 present the intergenerational findings.

We find that poor childhood health as proxied by hospitalization reduces the probability of achieving white-collar occupational status as an adult by 4.1 percentage points (pp), which corresponds to a 14.9 percent effect relative to the mean. The estimates for the likelihood of entering skilled or white-collar occupations, or the likelihood of attaining semi-skilled status and above are similar in magnitude although the effect sizes are smaller at 7.3 and 4.0 percent, respectively. Together, these findings indicate that the hospital patients were more likely than their siblings to end up in unskilled occupations in adulthood as a consequence of their poor health during childhood.

Given that the binary measure of hospitalization treats all causes of admission equally, we also consider a treatment variable that exploits the variation in severity across causes of admission (see Section 3.1 and appendix B) and is standardized on a 0 to 1 scale with higher values indicating more severe diagnoses. Among the hospital patients in the estimation sample, the mean value of the health deficiency index is 0.28 (s.d. 0.10), and representative causes of admission around this value include diseases of the hip or knee as well as rheumatism. A one-standard deviation increase in the health deficiency index decreases the likelihood of attaining higher occupational status as an adult by 1.6 to 5.8 percent. To fix ideas, a change in severity of this magnitude relative to the mean of the health deficiency index corresponds to being admitted for heart disease (morbus cordis) or sequela of diphtheria (diphtheric paralysis) as opposed to the conditions at the mean.

In addition to the effects on individuals' occupational status, we also find an impact of poor childhood health on intergenerational mobility. In comparison to their brothers, hospitalization increases the likelihood that a patient experienced downward occupational mobility relative to their father by 3.9 pp (13.9 percent). Likewise, a 1 s.d. (0.10) increase in the health deficiency index corresponds to a 1.1 pp (5.0 percent) increase in downward mobility. The effects on the probability of maintaining or rising in status relative to one's father are similar in magnitude, but the effect sizes are smaller relative to the mean, which is consistent with the high rates of upward mobility from unskilled jobs into skilled occupations during this time period (Long 2013).

5.1 Interpreting the magnitudes

To put the magnitude of our estimates for the effects of poor childhood health on long-run occupational outcomes into context we compare them to the association between father and son's status. We conduct this analysis on a sample of 2 million children age 0 to 11 in London linked from the 1881 to the 1911 census complete-count file. In Table A9, panel A scales the estimates from columns 1 to 3 of panel B in Table 2 to represent the impact of a 1 s.d. (0.10) change in the health deficiency index, while panel B presents results from estimating the following regressions:

$$\textit{Son's status}_i = \alpha + \beta \cdot \textit{Father's status}_i + \gamma X_i + \varepsilon_i \tag{2}$$

where i indexes individual, and *Son's status* and *Father's status* are one of three identical measures of status as adults: indicators for white collar occupations, skilled occupations and above, as well as semi-skilled occupations and above. X_i is a vector of individual characteristics that includes son's age and father's age fixed effects, standardized measures of first name frequency and an interaction of first name and surname frequency, indicators for a first born child, above median sibship size, and match quality (exact matches on first name, surname, and birth parish between censuses, or cleaned birth places). These regressions represent the closest approximation to an intergenerational wage elasticity that we can estimate given the variables in the census data. Having a high status father increases the likelihood of attaining high status as an adult by 18 to 25 percentage points.

To benchmark the magnitudes of our occupational status results, we first consider the effects of a 1 s.d. (0.10) change in the health deficiency index scaled by the estimates for the intergenerational transmission of occupational status. Recall that a change in the health deficiency index of this magnitude relative to the mean corresponds to suffering from diphtheric paralysis compared to a diseased knee or hip. Such a change offsets about 6.3 to 7.3 percent of the advantage of having a higher status father. Alternatively, if we scale the estimated coefficients on the hospitalization indicator from panel A of Table 2, we find that on average, health deficiencies in childhood offset 16.2 to 18.6 percent of the intergenerational status premium. Either way, the consequences of poor childhood health appear socially and economically relevant in London during the early-twentieth century.

We also consider benchmarks for the intergenerational mobility results based on the estimated degree of upward and downward occupational mobility in our sample and in the population. To do so, we construct occupational transition matrices for fathers and sons in our main estimation sample, and for the population of sons age 18 to 41 in 1911 linked to their fathers in 1881 (Table A10). Then, in Table A11 we sum the elements below the diagonal to obtain the share of

downward mobility in the population. The rate of downward mobility in our estimation sample is slightly higher than estimates for the population of England from 1881 to 1911 at 27.9 percent compared to 26.7 percent. When we scale our estimates for the effect of a 1 s.d. (0.10) increase in the health deficiency index on the chances of downward mobility, we find that it accounts for 5.0 to 5.2 percent of the overall degree of downward mobility in the sample and population, respectively. Similarly, the scaled effect for patients relative to their siblings accounts for 13.9 to 14.6 percent of overall downward mobility. The scaled effects of the patient indicator and the HDI on upward mobility are smaller at 5.4 to 5.5 percent, and 1.9 percent, respectively. Taken together, poor childhood health accounts for a meaningful but reasonable share of overall mobility in London, specifically, and England more generally at the turn of the twentieth century.

5.2 Mechanisms

Next, we consider potential mechanisms and mediators linking poor childhood health to long-run occupational outcomes and intergenerational mobility. We inquire whether poor childhood health limited opportunities for human capital accumulation, which in turn could have impeded success in the labor market. While we lack the data to conclude if the specific pathway is an adverse effect on cognitive ability, chronic health conditions that prevented regular school attendance, parental reinforcement of the health shock that reduced human capital investments, or a combination of these factors, we examine this mechanism using school attendance as an intermediate outcome. We interpret school attendance as broadly capturing the role of human capital accumulation. These results are presented in Table 3.

As the analysis of school attendance only requires a linkage from the hospital records to censuses during childhood, we can study the effects of poor childhood health on girls in addition to boys. Columns 1 and 2 display results from single sex samples where we compare male patients to their brothers and female patients to their sisters. Columns 3 to 5 then display results from larger samples that include both boys and girls and control for gender. First we pool together the single sex samples, then we restrict the sample to opposite sex comparisons within households, and finally we allow for either same sex and opposite sex comparisons.[23] As before, we present the results for the effect of hospitalization in panel A and the effect of the health deficiency index in panel B. Across the various specifications, the results are very similar, and hospitalization is associated with a 1.9 to 3.0 percentage point (2.5 to 4.1 percent) decrease in the probability of attending school 0 to 5 years after hospitalization. Similarly, a 1 s.d. (0.10) increase in the health

[23] A patient in the column 5 sample is compared to his or her sibling closest in age who is matched to the outcome-year census, regardless of gender. Since patients may only be matched to the closest sibling of one gender, the sample size in column 5 is smaller than the combined sample sizes of columns 3 and 4.

deficiency index leads to a 1.0 to 1.3 percent decrease in school attendance. Furthermore, it is notable that the effects of poor childhood health on schooling are modestly larger for girls than for boys. Thus, we conclude that the reduction in school attendance likely impeded human capital accumulation which is a plausible explanation for part of the decline in occupational status as adults for boys and potentially for girls.

In addition to the human capital channel documented in Table 3, we explore other mechanisms through heterogeneity analysis. Studies that utilize historical census data are typically unable to study heterogeneity to the same extent as those that use modern data, and our investigation is no exception. Nonetheless, we present sample splits that provide some suggestive evidence on how our results vary across relevant sub-populations. We examine three potential mediating channels: the severity of the health insult, the age at admission, and household socioeconomic status during childhood. Table A12 presents these results for the educational outcome (column 1), the three occupational outcomes (columns 2 to 4), and the two intergenerational outcomes (columns 5 and 6). It is important to note, however, that this evidence should be treated as suggestive only since due to relatively small sample sizes we cannot reject the null of equal coefficients at conventional levels of statistical significance in virtually all specification.

An advantage of using hospital records rather than intervention-specific health shocks to study childhood health is that the causes of admission to a hospital are wide-ranging, from conditions such as eczema that are not expected to yield detrimental effects, to those such as diphtheria that may cause severe consequences in the long-run. The health deficiency index introduced in Section 3.1 summarizes this variability, and thus in panel A of Table A12 we estimate separate effects for patients admitted for conditions with a health deficiency index above and below the median. The results for school attendance, white collar occupational status, and the two intergenerational mobility measures indicate that effects of poor health were more adverse among lower mortality admissions. These findings are consistent with positive selection among surviving patients admitted for high-mortality conditions relative to those with low-mortality conditions. Conversely, the magnitude of the estimates is larger for higher-mortality admissions when we consider effects on lower levels of occupational attainment (columns 3 and 4).

If earlier childhood experiences matter more than later ones then health shocks earlier in childhood could be more consequential than those experienced later. Thus, in panel B, we present results in which we estimate separate effects for children who were admitted to the hospital before age 5 and those that were admitted at ages 5 to 11.[24] Our point estimates are between 1.1 and 2 times larger for children admitted early rather than late in childhood, with the exception of

[24] 34.2 percent of patients in our main sample were admitted only at ages 0 to 4, 63.2 percent of patients were admitted only at ages 5 to 11, and 2.6 percent were admitted during both age ranges.

downward mobility for which the estimates are very similar. For the most part, however, the estimates are negative and statistically significant for both age groups. Thus, our findings not only provide qualitative support to literature on the importance of early childhood health, but also suggest that poor health later in childhood can likewise have detrimental effects on occupational status in the long run.

The final heterogeneity analysis, in panel C, explores the notion that a family's socioeconomic status may mediate the consequences of lower health capital during childhood (Currie and Hyson 1999). Wealthier parents who compensate for a child's poor health could provide additional resources to the child and minimize the potential negative consequences. On the other hand, if parents reinforce the initial health insult they will divert resources away from the affected child which magnifies the difference between a child observed in hospital records and their brothers. As patients from both high and low SES backgrounds have worse occupational outcomes in the long-run, we interpret our results as the direct negative effects of childhood health deficiencies bundled with any compensating or reinforcing behavior by parents.

6 Robustness

In Section 5 we documented the causal effects of poor childhood health on long-run occupational success and intergenerational mobility, and highlighted school attendance as an intermediate outcome and potential mechanism. Next, we conduct an extensive series of robustness checks to ensure that our results are stable and invariant to plausible alternative specifications and sample modifications. We address concerns that our results are biased downward by selective mortality or driven by outliers and examine the sensitivity of the results to changes in the criteria for inclusion in the sample and changes in the linking algorithm. The results are presented in Tables 4 and 5 for long-run outcomes with the occupational outcomes in panels A to C and the intergenerational outcomes in panels D and E. In each table we present our baseline estimates in the first column and show results using the hospitalized patient treatment indicator. In the appendix, we reproduce the same specifications using the health deficiency index (Tables A13 and A14) and present similar robustness analysis for school attendance (Table A15).

While we think of our estimates as capturing the effects of health deficiencies during childhood, they could also encompass poor health that persists into adulthood and impedes employment opportunities or productivity. Our main results may not be generalizable if they are driven by particular sub-groups of patients who suffered from severe or chronic health insults that resulted in a persistent scarring effect. Alternatively, sub-groups characterized by high mortality may be positively selected in the linked, longitudinal samples, resulting in downward bias to

our estimates. Either of these concerns apply to patients admitted to the hospital with severe conditions, as reflected by a high value of the health deficiency index (HDI), patients admitted as infants, and patients with recurring admissions. In separate estimation samples in columns 2, 4, and 5 of Table 4, we drop individuals admitted with health conditions in the top 10 percent of the HDI distribution (e.g. diphtheria or bronchopneumonia), individuals admitted at ages 0 or 1, and individuals admitted more than once. In column 3, we drop patients in the bottom 10 percent of the HDI distribution because mild health deficiencies are less likely to have long-run consequences and treating these individuals as having poor health would bias our estimates downward.[25] Another potential source of downward bias is within-household contagion and thus in column 6 we drop individuals admitted with contagious diseases. In each case, we also drop the siblings of the excluded hospitalized patients from the robustness exercises since they would not contribute to identification in the specifications with sibling fixed effects. Across these specifications with varying sample restrictions, our coefficients remain very similar and if anything, in many instances the negative effects are slightly larger in magnitude, which suggests that our main estimates are biased downward due to positive selection of surviving patients, the inclusion of mild health deficiencies with no long-term consequences in the treatment, and within-household contagion. Ruling out that our results are driven by poor health during infancy, when mortality was very high, also supports our interpretation that poor health during childhood has negative long-run consequences.

The next set of robustness checks addresses different sources of sample selection bias and concerns about external validity. In our main specification, we compare a male hospital patient to the brother closest in age among those linked between censuses. Our sibling fixed effects estimates could be biased by the comparison to a specific non-hospitalized sibling to the extent that there are negative spillovers from the patient's hospitalization to siblings close in age. Thus, in column 2 of Table 5 we add to the comparison group all linked brothers within 8 years of the patient's age. Sample selection bias could also arise from the set of households included in the analysis, given that in our main specification we restrict attention to households with only one male patient linked from the hospital records to censuses during childhood and adulthood. In column 3 we add families with multiple hospital patients to demonstrate that our results are unaffected by this sample restriction.

Another potential concern with our estimates is bias due to differential selection of patients by distance traveled to the hospital. The general hospitals (Barts and Guy's) and the children's

[25]Moreover, the most common causes of admission in the bottom 10 percent of the HDI distribution were club foot (talipes) and hare lip, conditions for which surgical treatment was available. In order to treat club foot and other developmental deformities that involved the shortening of tendons, surgeons performed a tenotomy, which was a routine procedure by the end of the nineteenth century. Similarly, operations to repair hare lip was considered relatively simple and safe by 1860 (Lomax 1996).

hospital (GOSH) in our sample admitted 15, 12, and 31 percent of their patients from outside the County of London, respectively. Providing care to children from outside of London was especially common at GOSH, which specialized in the treatment of rare childhood medical conditions. As parents faced a higher cost of bringing the child to the hospital the further they traveled, patients residing outside of London could have had greater health deficiencies or conversely, particularly good unobservable characteristics that made it worthwhile to invest in hospital care. Furthermore, individuals residing outside the County of London when admitted to the hospital are 4.2 percentage points less likely to be matched to any census record, which indicates that linked individuals may be selective. However, column 4 shows that our results are robust to restricting the sample to patients residing in the County of London at the time of hospitalization, easing concerns about selection by distance travelled. Finally, for Barts Hospital and GOSH we observe the universe of admissions seen by physicians and surgeons, while for Guy's we only obtained the records of patients seen by physicians, and thus the latter may be unrepresentative of hospital admissions.[26] Thus, in column 5 we show that the main results are robust to dropping households with patients admitted to Guy's Hospital. Taken together, these robustness exercises address external validity concerns and selection bias due the set of siblings, households, residential locations, or hospitals included in the main analysis sample.

An issue that arises with the use of automated linking methods and imperfect historical data is that linked samples are likely to contain false positive matches (Bailey et al. 2017). When determining the strictness of the linking criteria, the researcher faces a trade-off between the number of false positives and the sample size, and decreasing the latter reduces precision and external validity as linked observations become highly selective. Our application imposes significant demands on the automated linking methods as we require three separate linkages: patients from hospital records to childhood census as well as patients and siblings between censuses in childhood and adulthood. For our preferred estimates, we allow matched records to have names with Jaro-Winkler distances up to 0.2, other records with similar names in neighboring birth cohorts, and ages that differ by up to 3 years. Our approach is supported by Abramitzky et al. (2019) who have shown in other contexts that coefficients of interest are typically stable across specifications with stricter versus lenient match criteria. While we lack "ground truth" data to assess the extent of false positive, we can illustrate the consistency of our results when we impose stricter criteria that reduce the likelihood of false matches. We also demonstrate the stability of our results when we relax the thresholds for linked records.

We begin the linkage diagnostics by probing for bias from false positive matches due to first names or surnames that differ in their Jaro-Winkler distance scores. In Figure A9 we plot the co-

[26]Inpatient hospital admissions in nineteenth-century London were categorized as physician or surgeon patients. Among cohorts in our samples, physician patients accounted for 35 percent of inpatients at Barts Hospital.

efficient on the patient indicator across the six dependent variables as we vary the Jaro-Winkler distance threshold by increments of 0.025 from 0 to 0.2 in the main specification. Our results are qualitatively similar when we impose the more restrictive specifications, though we lose precision and our estimates shrink somewhat for the white collar and skilled-plus outcomes when we restrict to exactly matched names. In addition, they are very stable for the intergenerational and school attendance outcomes. In the presence of misreported ages in the census, the probability of a false match is also increasing in the number of records with similar names in neighboring birth cohorts. False matches would introduce measurement error in the dependent variable and reduce the precision of our estimates. In the main specification the maximum number of records with similar names in the outcome-year census is 20 and a similar name is defined as differing in the Jaro-Winkler distance score by less than 0.10. In Figure A10 we allow the number of similar names to vary between 4 and 1000, and restrict the definition of similar names to include records with a difference in the Jaro-Winkler distance of less than 0.05 or 0.10.[27] Our estimates are comparable across the various permutations though sometimes, as expected, we lose some precision in specifications with few similar names due to much reduced sample sizes and with many similar names due to a much higher likelihood of false matches.

Returning to Table 5, we present two additional robustness checks in columns 6 and 7 that tweak the linking procedure to address concerns about false positives and sample selection bias. In our main specification, we follow the standard approach of automated linking methods by not using place of residence as a linking variable due to concerns about endogenous residential choice. However, in the absence of birth place information in the hospital records, we can only link individuals to the census using name and age, and thus to break ties we first prioritize records with the same county of residence, followed by records with the same district or parish of residence. As it was common in this time period for households to move frequently between registration districts or parishes, the latter step may introduce false positive matches and could bias our sample towards stayers. Cross-county moves were less common and so prioritizing records that match on county of residence may not bias our results, but the possibility of false positives remains. Thus, we impose an additional restriction that individuals linked between the hospital and census records are uniquely identified within the county of residence at enumeration to avoid biasing the sample towards non-movers (column 6). Then, we further restrict the sample to records with the same county of residence in the hospital records and the census to balance the trade-off between restricting to non-movers and the potential for false matches across counties (column 7). In both cases the results are consistent with the main estimates, indicating that the way we incorporate place of residence in the matching procedure should not introduce bias.

[27] When we restrict the sample to records with fewer than 4 similar names in neighboring birth cohorts, the sample size decreases significantly and becomes characterized by rare names.

The robustness exercises demonstrate that our main estimates for the effect of poor childhood health on school attendance, occupational success, and intergenerational mobility are stable across a variety of specification choices and remain statistically significant in the vast majority of permutations. In particular, we show that the results are robust to changes in the definition of the treatment, the use of alternative sample selection criteria, and variation in the strictness of the matching criteria. We thus conclude that the results are unlikely to be driven by our preferred choice of specification in the main analysis, or by false positive matches.

7 Conclusion

An extensive body of research has documented long-run and intergenerational consequences of prenatal health shocks, but much less is known about the role of health capital during childhood. To the extent that health matters post-infancy, it may be an important piece of the puzzle missing from explanations for socioeconomic inequality within and across generations. In this paper, we show that health deficiencies proxied by hospital admissions at ages 0 to 11 affect school enrollment and long-run occupational success, and contribute to explaining the intergenerational elasticity between fathers and sons.

To estimate the impact of health deficiencies during childhood, we link records of individual hospital admissions that occurred between 1870 and 1902 to longitudinal samples of the Census of England from 1881 to 1911. Then, we implement a sibling fixed effects identification strategy that contrasts patients with their siblings who lived in the same household during childhood but were not hospitalized as far as we can observe in the surviving records. We find that compared to their brothers, hospitalized males were 14.9 percent less likely to work in white collar occupations and 22.7 percent more likely to work in unskilled occupations as adults. In the first case, the effect size offsets 16.2 percent of the advantage of having a father in while collar occupation. Consistent with these results, we also find that these patients were 13.9 percent more likely than their brothers to experience downward occupational mobility.

Our results are robust to a variety of specification checks including issues related to selective mortality, sample selection choices, or changes to the matching algorithms. In considering explanations for the estimated effects, we show that hospitalized boys were 2.5 percent less likely to attend school compared to their non-hospitalized brothers in the 5-year period after the hospital admission. Although we cannot observe long-run outcomes for girls due to name changes at marriage, we find that the effects of health capital on school attendance are modestly larger for girls than for boys at 4.1 percent. This implies that as adults these females potentially experienced even more detrimental socioeconomic outcomes.

Today, the rapid urbanization as well changes in the disease environment and mortality trends taking place in developing countries are arguably similar to the historical experience of late-nineteenth century England (Mercer 2014), notwithstanding differences in public health infrastructure and the availability of antibiotics. As was the case historically, serious and unexpected illnesses remain a significant threat to the economic opportunities of households in these counties, especially in the absence of formal health and disability insurance schemes (Gertler and Gruber 2002). Although quantifying the exact mitigating role of a social safety net is beyond the scope of this paper, a safety net can smooth health shocks and thus contribute to productivity and welfare gains for children upon reaching adulthood, and potentially for the next generation as well. We view understanding the exact role that various channels play in driving the estimated effects in both historical and modern contexts as a fruitful avenue of future research.

References

Abramitzky, Ran, Leah Platt Boustan, and Katherine Eriksson. 2012. Europe's Tired, Poor, Huddled Masses: Self-Selection and Economic Outcomes in the Age of Mass Migration. *American Economic Review* 102, no. 5:1832–1856.

Abramitzky, Ran, Leah Platt Boustan, Katherine Eriksson, James J. Feigenbaum, and Santiago Pérez. 2019. Automated Linking of Historical Data. *National Bureau of Economic Research Working Paper Series* No. 25825.

Abramitzky, Ran, Adeline Delavande, and Luis Vasconcelos. 2011. Marrying Up: The Role of Sex Ratio in Assortative Matching. *American Economic Journal: Applied Economics* 3, no. 3:124–157.

Adhvaryu, Achyuta, Steven Bednar, Teresa Molina, Quynh Nguyen, and Anant Nyshadham. 2019. When It Rains It Pours: The Long-run Economic Impacts of Salt Iodization in the United States. *The Review of Economics and Statistics* :1–45.

Almond, Douglas and Janet Currie. 2011a. Human Capital Development Before Age Five. In *Handbook of labor economics*, vol. 4B, eds. Orley Ashenfelter and David Card. Amsterdam: Elsevier, 1315–1486.

———. 2011b. Killing Me Softly: The Fetal Origins Hypothesis. *Journal of Economic Perspectives* 25, no. 3:153–172.

Almond, Douglas, Janet Currie, and Valentina Duque. 2018. Childhood Circumstances and Adult Outcomes: Act II. *Journal of Economic Literature* 56, no. 4:1360–1446.

Arthi, Vellore and Eric B. Schneider. 2017. Infant Feeding and Cohort Health: Evidence from the London Foundling Hospital. Working Paper.

Bailey, Martha, Connor Cole, Morgan Henderson, and Catherine Massey. 2017. How Well Do Automated Methods Perform in Historical Samples? Evidence from New Ground Truth. *National Bureau of Economic Research Working Paper Series* No. w24019.

Baird, Sarah, Joan Hamory Hicks, Michael Kremer, and Edward Miguel. 2016. Worms at Work: Long-run Impacts of Child Health Gains. *Quarterly Journal of Economics* 131, no. 4:1637–1680.

Barreca, Alan I. 2010. The Long-Term Economic Impact of in Utero and Postnatal Exposure to Malaria. *Journal of Human Resources* 45, no. 4:865–892.

Beach, Brian, Joseph Ferrie, Martin Saavedra, and Werner Troesken. 2016. Typhoid Fever, Water Quality, and Human Capital Formation. *The Journal of Economic History* 76, no. 1:41–75.

Beach, Brian, Joseph P Ferrie, and Martin H Saavedra. 2018. Fetal Shock or Selection? The 1918 Influenza Pandemic and Human Capital Development. *National Bureau of Economic Research Working Paper Series* No. 24725.

Bhalotra, Sonia and Atheendar S. Venkataramani. 2012. Shadows of the Captain of the Men of Death: Early Life Health Interventions, Human Capital Investments, and Institutions. *Working Paper* .

Bleakley, Hoyt. 2007. Disease and Development: Evidence from Hookworm Eradication in the American South. *The Quarterly Journal of Economics* 122, no. 1:73–117.

———. 2010. Malaria Eradication in the Americas: A Retrospective Analysis of Childhood Exposure. *American Economic Journal: Applied Economics* 2, no. 2:1–45.

Brown, David W, Amanda E Kowalski, and Ithai Z Lurie. 2019. Long-Term Impacts of Childhood Medicaid Expansions on Outcomes in Adulthood. *The Review of Economic Studies* forthcoming.

Brunton, Deborah. 2004. *Medicine transformed: Health, disease and society in Europe, 1800-1930.* Manchester: Manchester University Press.

Bütikofer, Aline, Katrine V Løken, and Kjell G. Salvanes. 2019. Infant Health Care and Long-Term Outcomes. *The Review of Economics and Statistics* :1–14.

Bütikofer, Aline and Kjell G. Salvanes. 2019. Disease Control and Inequality Reduction: Evidence from a Tuberculosis Testing and Vaccination Campaign. *Review of Economic Studies* .

Carpenter, Mary Wilson. 2010. *Health, medicine, and society in Victorian England.* Santa Barbara, Calif.: Praeger.

Case, Anne, Angela Fertig, and Christina Paxson. 2005. The lasting impact of childhood health and circumstance. *Journal of Health Economics* 24, no. 2:365 – 389.

Case, Anne, Darren Lubotsky, and Christina Paxson. 2002. Economic Status and Health in Childhood: The Origins of the Gradient. *American Economic Review* 92, no. 5:1308–1334.

Case, Anne and Christina Paxson. 2008. Stature and Status: Height, Ability, and Labor Market Outcomes. *Journal of Political Economy* 116, no. 3:499–532.

———. 2010. Causes and Consequences of Early-Life Health. *Demography* 47, no. 1:S65–S85.

Chatto and Windus. 1897. Sixty years of healing. *Belgravia: A London magazine* 93:203–233.

Christen, Peter. 2012. *Data Matching: Concepts and Techniques for Record Linkage, Entity Resolution, and Duplicate Detection.* Heidelberg: Springer.

Clark, Gregory. 2014. *The Son Also Rises: Surnames and the History of Social Mobility.* Princeton, NJ: Princeton University Press.

Costa, Dora L. 2015. Health and the Economy in the United States from 1750 to the Present. *Journal of Economic Literature* 53, no. 3:503–570.

Cullis, P. S. and D. G. Young. 2013. Nineteenth century mortality at the Royal Hospital for Sick Children, Glasgow. *Scottish Medical Journal* 58, no. 3:182–187.

Currie, Janet. 2009. Healthy, Wealthy, and Wise: Socioeconomic Status, Poor Health in Childhood, and Human Capital Development. *Journal of Economic Literature* 47, no. 1:87–122.

Currie, Janet and Rosemary Hyson. 1999. Is the Impact of Health Shocks Cushioned by Socioeconomic Status? The Case of Low Birthweight. *American Economic Review* 89, no. 2:245–250.

Currie, Janet, Mark Stabile, Phongsack Manivong, and Leslie L. Roos. 2010. Child Health and Young Adult Outcomes. *Journal of Human Resources* 45, no. 3:517–548.

Currie, Janet and Tom Vogl. 2013. Early-Life Health and Adult Circumstance in Developing Countries. *Annual Review of Economics* 5, no. 1:1–36.

Cutler, David, Angus Deaton, and Adriana Lleras-Muney. 2006. The Determinants of Mortality. *Journal of Economic Perspectives* 20, no. 3:97–120.

Doyle, Shane, Felix Meier zu Selhausen, and Jacob Weisdorf. 2019. The Blessings of Medicine? Patient Characteristics and Health Outcomes in a Ugandan Mission Hospital, 1908-1970. Working Paper.

Feigenbaum, James J. 2016. A Machine Learning Approach to Census Record Linking. Working Paper.

Ferrie, Joseph P. 1996. A New Sample of Males Linked from the Public Use Micro Sample of the 1850 U.S. Federal Census of Population to the 1860 U.S. Federal Census Manuscript Schedules. *Historical Methods* 29:141–156.

Floud, Roderick, Robert W. Fogel, Bernard Harris, and Sok Chul Hong. 2011. *The Changing Body : Health, Nutrition, and Human Development in the Western World since 1700.* Cambridge: Cambridge University Press.

Fogel, Robert William. 2004. *The Escape from Hunger and Premature Death, 1700-2100.* Cambridge: Cambridge University Press.

Franklin, Alfred White. 1964. Children's Hospitals. In *The evolution of hospitals in Britain*, ed. F. N. L. Poynter. London: Pitman, 103–121.

Gensowski, Miriam, Torben Heien Nielsen, Nete Munk Nielsen, Maya Rossin-Slater, and Miriam Wüst. 2018. Childhood Health Shocks, Comparative Advantage, and Long-Term Outcomes: Evidence from the Last Danish Polio Epidemic. *National Bureau of Economic Research Working Paper Series* No. 24753.

Gertler, Paul and Jonathan Gruber. 2002. Insuring Consumption against Illness. *American Economic Review* 92, no. 1:51–70.

Goodman-Bacon, Andrew. 2017. The Long-Run Effects of Childhood Insurance Coverage: Medicaid Implementation, Adult Health and Labor Market Outcomes. Working Paper.

Gould, Eric D., Victor Lavy, and M. Daniele Paserman. 2011. Sixty Years after the Magic Carpet Ride: The Long-Run Effect of the Early Childhood Environment on Social and Economic Outcomes. *The Review of Economic Studies* 78, no. 3:938–973.

Hardy, Anne. 2001. *Health and Medicine in Britain since 1860.* New York: Palgrave.

Hawkins, Sue and Andrea Tanner. 2013. Myth, Marketing, and Medicine: Life in British Children's Hospitals, 1850-1914. In *Hospital life: Theory and practice from the medieval to the modern*, eds. Laurinda Abreu and Sally Sheard. Bern, Switzerland: Peter Lang, 209–236.

Higgs, Michelle. 2009. *Life in the Victorian hospital.* Stroud, Gloucestershire: History Press.

Hjort, Jonas, Mikkel Sølvsten, and Miriam Wüst. 2017. Universal Investment in Infants and Long-Run Health: Evidence from Denmark's 1937 Home Visiting Program. *American Economic Journal: Applied Economics* 9, no. 4:78–104.

Hoehn-Velasco, Lauren. 2019. The Long-term Impact of Preventative Public Health Programs. Working Paper.

Hong, Sok Chul. 2007. The Burden of Early Exposure to Malaria in the United States, 1850-1860: Malnutrition and Immune Disorders. *The Journal of Economic History* 67, no. 04:1001–1035.

———. 2013. Malaria: An Early Indicator of Later Disease and Work Level. *Journal of Health Economics* 32, no. 3:612–632.

Horrell, Sara, Jane Humphries, and Hans-Joachim Voth. 2001. Destined for Deprivation: Human Capital Formation and Intergenerational Poverty in Nineteenth-Century England. *Explorations in Economic History* 38, no. 3:339–365.

Jayachandran, Seema, Adriana Lleras-Muney, and Kimberly V. Smith. 2010. Modern Medicine and the Twentieth Century Decline in Mortality: Evidence on the Impact of Sulfa Drugs. *American Economic Journal: Applied Economics* 2, no. 2:118–146.

Karbownik, Krzysztof and Anthony Wray. 2015. Childhood health and long-run economic opportunity in Victorian England. Working Paper.

Kingston University. 2010. *HHARP: The Historic Hospital Admission Records Project.* Kingston upon Thames, Surrey: Kingston University.

Lazuka, Volha. 2019. Infant Health and Later-Life Labor Market Outcomes: Evidence from the Introduction of Sulpha Antibiotics in Sweden. *Journal of Human Resources* Forthcomin.

Lomax, Elizabeth M. R. 1996. *Small and Special: The Development of Hospitals for Children in Victorian Britain.* London: Wellcome Institute for the History of Medicine.

Long, Jason. 2013. The Surprising Social Mobility of Victorian Britain. *European Review of Economic History* 17, no. 1:1–23.

Long, Jason and Joseph Ferrie. 2013. Intergenerational Occupational Mobility in Great Britain and the United States since 1850. *American Economic Review* 103, no. 4:1109–1137.

McKeown, Thomas. 1976. *The Modern Rise of Population.* London: Edward Arnold.

Mercer, Alexander. 2014. *Infections, chronic disease, and the epidemiological transition : a new perspective.* Rochester, NY: University of Rochester Press.

Mill, Roy and Luke C. D. Stein. 2016. Race, Skin Color, and Economic Outcomes in Early Twentieth-Century America. Working Paper.

Miller, Douglas L, Na'ama Shenhav, and Michel Z Grosz. 2019. Selection into Identification in Fixed Effects Models, with Application to Head Start. *National Bureau of Economic Research Working Paper Series* No. 26174.

Mokyr, Joel. 2000. Why "More Work for Mother?" Knowledge and Household Behavior, 1870-1945. *The Journal of Economic History* 60, no. 1:1–41.

Nix, Emily and Nancy Qian. 2015. The Fluidity of Race: "Passing" in the United States, 1880-1940. Working Paper.

Olivetti, Claudia and Daniele Paserman. 2015. In the Name of the Son (and the Daughter): Intergenerational Mobility in the United States, 1850-1940. *American Economic Review* 105, no. 8:1–31.

Parman, John. 2015a. Childhood health and human capital: New evidence from genetic brothers in arms. *Journal of Economic History* 75, no. 1:30–64.

———. 2015b. Childhood Health and Sibling Outcomes: Nurture Reinforcing Nature During the 1918 Influenza Pandemic. *Explorations in Economic History* 58:22–43.

Pérez, Santiago. 2019. Intergenerational Occupational Mobility across Three Continents. *The Journal of Economic History* 79, no. 2:383–416.

Ruggles, Steven. 2014. Big microdata for population research. *Demography* 51, no. 1:287–297.

Saavedra, Martin. 2017. Early-life disease exposure and occupational status: The impact of yellow fever during the 19th century. *Explorations in Economic History* 64:62–81.

Schürer, K and Joe Day. 2019. Migration to London and the development of the north–south divide, 1851–1911. *Social History* 44, no. 1:26–56.

Schwandt, Hannes. 2018. The Lasting Legacy of Seasonal Influenza: In-Utero Exposure and Labor Market Outcomes. Working Paper.

Smith, James P. 2009. The Impact of Childhood Health on Adult Labor Market Outcomes. *Review of Economics and Statistics* 91, no. 3:478–489.

Szreter, Simon. 2005. *Health and Wealth.* Rochester, NY: University of Rochester Press.

UK Data Archive. 2014. *International Census Microdata Project [Electronic Dataset].* Colchester, UK: University of Essex.

van Leeuwen, Marco H. D. and Ineke Maas. 2011. *Hisclass: A Historical International Social Class Scheme.* Leuven, Belgium: Leuven University Press.

van Leeuwen, Marco H. D., Ineke Maas, and Andrew Miles. 2002. *Hisco: Historical International Standard Classification of Occupations.* Leuven, Belgium: Leuven University Press.

Venkataramani, Atheendar S. 2012. Early Life Exposure to Malaria and Cognition in Adulthood: Evidence from Mexico. *Journal of Health Economics* 31, no. 5:767–780.

Wohl, Anthony S. 1983. *Endangered Lives: Public Health in Victorian Britain.* Cambridge, MA: Harvard University Press.

Worboys, Michael. 2000. *Spreading Germs: Diseases, Theories, and Medical Practice in Britain, 1865-1900.* New York: Cambridge University Press.

Zhang, Xiaohan. 2014. Disease and Development of the Next Generation. Working Paper.

8 Tables

Table 1: Linkage rates from hospital records to childhood census

	Census year linked to hospital records			
	(1) 1881	(2) 1891	(3) 1901	(4) Any
	Panel A: Hospital to childhood census linkage			
No match	1,016	1,243	1,779	983
	(0.064)	(0.054)	(0.167)	(0.035)
Multiple matches	10,603	15,318	6,255	17,702
	(0.667)	(0.662)	(0.586)	(0.622)
Unique match	4,284	6,562	2,635	9,763
	(0.269)	(0.284)	(0.247)	(0.343)
Sibling present	4,007	5,945	2,240	9,171
	(0.252)	(0.257)	(0.210)	(0.322)
Long-run censuses	1901, 1911	1901, 1911	1911	1901, 1911
	Panel B: Linkage to census in adulthood			
Patient matched	2,968	3,749	1,380	6,238
	(0.187)	(0.162)	(0.129)	(0.219)
Patient and sibling	1,195	1,572	362	3,105
	(0.075)	(0.068)	(0.034)	(0.109)
In final sample	970	1,082	269	2,320
	(0.061)	(0.047)	(0.025)	(0.082)
Total admissions	15,903	23,123	10,669	28,448

Notes: Panel A of this table presents sample sizes and linkage rates (in parentheses) from inpatient hospital admission records to the 1881, 1891, and 1901 census (columns 1 to 3), as well from pooling together linkages across the three censuses (column 4). "Total admissions" represents the number of admissions that we attempt to match to the census in each case. The sample includes all patients from the 1870 to 1890 birth cohorts who were admitted at ages 0 to 11 between 1870 and 1902 no more than 10 years before or after enumeration in the census. In the top panel we first show the number of admissions with either multiple matches, no match or a unique match in each census. Then, we indicate the subset of the unique matches for whom we also match a brother. Panel B of the table shows the number of patients who were also linked to a census during adulthood, the subset of these individuals for whom a sibling was also matched, and the final sample of observations.

Table 2: Occupational status and intergenerational mobility

	(1) White collar	(2) Skilled +	(3) Semi-skilled +	(4) Class ↘	(5) Class ↗
			Panel A: Effects of hospital admission		
Patient	−0.041***	−0.040**	−0.034***	0.039**	−0.039**
	(0.014)	(0.016)	(0.011)	(0.015)	(0.015)
% effect	14.9	7.3	4.0	13.9	5.5
			Panel B: Effects of health deficiency index		
Health deficiency index	−0.122***	−0.120**	−0.104***	0.108**	−0.106**
	(0.044)	(0.053)	(0.037)	(0.053)	(0.053)
% effect (σ)	5.8	2.8	1.6	5.0	1.9
Mean of Y	0.273	0.544	0.850	0.279	0.721
N	3,996	3,996	3,996	2,606	2,606

Notes: The dependent variables in columns 1 to 3 are indicators of occupational status: the probability of entering a white-collar occupation; the probability of entering a white-collar or skilled occupation; and the probability of entering a white-collar, skilled, or semi-skilled occupation. In columns 4 and 5, the dependent variables are measures of intergenerational occupational mobility: the probability of having a lower occupational status than one's father, and the likelihood of the same or higher occupational status than one's father, respectively. Each regression includes age-by-outcome census year and birth-order fixed effects, an indicator for the older sibling in each pair of observations from a household, standardized measures of first name frequency, an interaction of first name and surname frequency, and indicators for match quality (exact matches on first name, surname, and birth parish between censuses, or cleaned birth places). The specifications with the intergenerational outcomes restrict the sample to cases where fathers and sons were less than 20 years apart in age at the time of census enumeration when occupational outcomes were observed. Standard errors are clustered by childhood household.

Point estimates marked ***, **, and * are statistically significant at the 1, 5, and 10 percent levels, respectively.

Table 3: School attendance

	Single sex samples		Both sexes in sample		
	(1) Males	(2) Females	(3) Same sex	(4) Opposite sex	(5) Mixed
	Panel A: Effects of hospital admission				
Patient	−0.019*	−0.030**	−0.024***	−0.019**	−0.020***
	(0.011)	(0.012)	(0.008)	(0.008)	(0.006)
% effect	2.5	4.1	3.2	2.5	2.7
	Panel B: Effects of health deficiency index				
Health deficiency index	−0.068*	−0.077*	−0.072***	−0.059**	−0.061***
	(0.038)	(0.040)	(0.027)	(0.029)	(0.020)
% effect (σ)	1.2	1.3	1.2	1.0	1.1
Mean of Y	0.750	0.736	0.743	0.741	0.745
N	1,530	1,510	3,040	2,830	4,832

Notes: In all columns, the dependent variable is an indicator for school attendance, which is proxied by the recording of "scholar" as an individual's occupation in the 1881 or 1891 Census of England. In each regression, the sample includes children aged 5 to 10 at the time of census enumeration. In panel A, the explanatory variable of interest is an indicator for hospital admission, and in panel B, it is the health deficiency index. Each regression includes age-by-census year and birth-order fixed effects, an indicator for the older sibling in each pair of observations from a household, standardized measures of first name frequency, an interaction of first name and surname frequency, and indicators for match quality (exact matches on first name and surname). In column 3, we pool together the samples in columns 1 and 2, and make same-gender comparisons within households. Column 4 is also based on a pooled sample, but makes opposite-gender comparisons within households, and thus also includes a control for gender. Column 5 is based on a larger sample that allows for both same-gender and opposite-gender comparisons within households, and also includes a control for gender. Standard errors are clustered by childhood household.

Point estimates marked ***, **, and * are statistically significant at the 1, 5, and 10 percent levels, respectively.

Table 4: Long-run outcomes: Robustness to selective mortality

	(1) Baseline Estimate	(2) Drop high mortality	(3) Drop low mortality	(4) Drop infant admission	(5) Drop multiple admissions	(6) Drop contagious
			Panel A: Effects on P(White collar)			
Patient	−0.041***	−0.045***	−0.046***	−0.044***	−0.048***	−0.047***
	(0.014)	(0.014)	(0.014)	(0.014)	(0.014)	(0.015)
Mean of Y	0.273	0.270	0.279	0.276	0.268	0.275
			Panel B: Effects on P(Skilled +)			
Patient	−0.040**	−0.046***	−0.043***	−0.031*	−0.037**	−0.041**
	(0.016)	(0.016)	(0.017)	(0.016)	(0.016)	(0.017)
Mean of Y	0.544	0.543	0.547	0.550	0.539	0.542
			Panel C: Effects on P(Semi-skilled +)			
Patient	−0.034***	−0.035***	−0.033***	−0.034***	−0.034***	−0.043***
	(0.011)	(0.011)	(0.011)	(0.012)	(0.012)	(0.012)
Mean of Y	0.850	0.848	0.853	0.851	0.852	0.850
N	3,996	3,596	3,594	3,628	3,528	3,282
			Panel D: Effects on P(Class ↘)			
Patient	0.039**	0.039**	0.038**	0.035**	0.040**	0.034**
	(0.015)	(0.016)	(0.017)	(0.016)	(0.016)	(0.017)
Mean of Y	0.279	0.273	0.281	0.278	0.279	0.283
			Panel E: Effects on P(Class ↗)			
Patient	−0.039**	−0.041***	−0.038**	−0.034**	−0.038**	−0.038**
	(0.015)	(0.016)	(0.017)	(0.016)	(0.016)	(0.017)
Mean of Y	0.721	0.728	0.719	0.721	0.721	0.718
N	2,606	2,338	2,330	2,356	2,310	2,126

Notes: Each cell displays the coefficient for the hospitalization indicator from a separate regression. The dependent variables in panels A to E mirror those in columns 1 to 5 of Table 2, respectively. Column 1 reproduces the estimates in panel A of Table 2. Columns 2 and 3 drop households with patients admitted for conditions in the upper and lower deciles of the health deficiency index distribution, respectively. Column 4 excludes households with patients aged 0 or 1 at admission, column 5 drops households with patients admitted multiple times, and column 6 leaves out households with patients admitted for contagious illnesses. See Table 2 for a description of control variables. Standard errors are clustered by childhood household.

Point estimates marked ***, **, and * are statistically significant at the 1, 5, and 10 percent levels, respectively.

Table 5: Long-run outcomes: Robustness to sample selection

	(1) Baseline Estimate	(2) Add multiple siblings	(3) Add multiple patient hhlds.	(4) County of London only	(5) Drop Guy's Hospital	(6) Unique within census county	(7) Hospital-census county match
				Panel A: Effects on P(White collar)			
Patient	−0.041***	−0.037***	−0.036***	−0.043***	−0.040***	−0.043***	−0.054***
	(0.014)	(0.013)	(0.013)	(0.015)	(0.014)	(0.014)	(0.017)
Mean of Y	0.273	0.274	0.274	0.281	0.278	0.277	0.287
				Panel B: Effects on P(Skilled +)			
Patient	−0.040**	−0.032**	−0.030**	−0.041**	−0.047***	−0.037**	−0.056***
	(0.016)	(0.014)	(0.014)	(0.017)	(0.016)	(0.016)	(0.019)
Mean of Y	0.544	0.543	0.543	0.551	0.553	0.544	0.554
				Panel C: Effects on P(Semi-skilled +)			
Patient	−0.034***	−0.026**	−0.027***	−0.037***	−0.034***	−0.033***	−0.028**
	(0.011)	(0.010)	(0.010)	(0.012)	(0.012)	(0.012)	(0.014)
Mean of Y	0.850	0.848	0.849	0.860	0.857	0.850	0.861
N	3,996	4,828	4,870	3,270	3,596	3,578	2,568
				Panel D: Effects on P(Class ↘)			
Patient	0.039**	0.030**	0.028**	0.031*	0.040**	0.035**	0.033*
	(0.015)	(0.013)	(0.013)	(0.017)	(0.016)	(0.016)	(0.019)
Mean of Y	0.279	0.280	0.282	0.281	0.278	0.278	0.276
				Panel E: Effects on P(Class ↗)			
Patient	−0.039**	−0.030**	−0.028**	−0.031*	−0.040**	−0.035**	−0.034*
	(0.015)	(0.013)	(0.013)	(0.017)	(0.016)	(0.016)	(0.019)
Mean of Y	0.721	0.720	0.718	0.720	0.722	0.722	0.724
N	2,606	4,021	4,053	2,198	2,334	2,292	1,700

Notes: Each cell displays the coefficient for the hospitalization indicator from a separate regression. The dependent variables in panels A to E mirror those in columns 1 to 5 of Table 2, respectively. Column 1 reproduces the estimates in panel A of Table 2. Column 2 compares patients to all male siblings linked between censuses instead of restricting the comparison to the male sibling closest in age. Column 3 adds households with multiple patients, column 4 restricts the sample to patients and siblings residing in the Greater London area during childhood, and column 5 drops households with patients admitted to Guy's Hospital. Column 6 restricts the sample to households with patients whose name and age combinations are unique within their county of residence in the childhood census to which they are linked. Column 7 further restricts to individuals residing in the same county at the time of hospitalization and census enumeration. See Table 2 for a description of control variables. Standard errors are clustered by childhood household.

Point estimates marked ***, **, and * are statistically significant at the 1, 5, and 10 percent levels, respectively.

For Online Publication: Online Appendix

A Data Linking Procedure

In this section we describe the linking algorithms used to match inpatient admission records to census records and to match individuals across censuses. We explain the census-to-census linkage in detail as it is the most basic procedure and it is the basis for the hospital-to-census linkage with some modifications.

A.1 Census-to-census linkages

We use the complete count data for the Census of England from the I-CeM project to create the following linked samples:

1. 1881 to 1901
2. 1891 to 1901
3. 1881 to 1911
4. 1891 to 1911
5. 1901 to 1911

In each case, we start with all males aged 0 to 21 in the base year census. We exclude females due to name changes at marriage which prevent matching based on maiden surname. Census-to-census linkages are based on time-invariant characteristics such as first name, surname, birth year, and county of birth[28]. We begin by separating given names into first and middle names, and then standardize diminutives and common nicknames of first names to their proper equivalents. We follow the procedure in Parman (2015b) and construct the Phonex codes for the first and last names in each data set, which enables us to allow for differences in the spelling of phonetically similar names across data sets that might arise from factors such as typographical errors.[29] Prior to the implementation of the matching algorithm, we perform a "blocking" step in which the two data sets are joined using four blocking variables: the Phonex code of the first and last names, age in years when enumerated in the later census, and county of birth (Christen 2012).

[28] We choose not to match on birth parish during the initial step given that the variable is a non-standardized text string and parish boundaries changed significantly over the time period of study

[29] See Nix and Qian (2015) and the supplemental materials to Parman (2015b) for discussions of the Phonex algorithm.

The linkage procedure draws on elements of the methods pioneered by Ferrie (1996) and utilized by Abramitzky et al. (2012), modifications developed by Feigenbaum (2016) and Mill and Stein (2016), and recommendations made by Bailey et al. (2017).[30] It proceeds as follows:

1. Re-code all births in the counties of Kent, Middlesex and Surrey as births in "London" to account for changes in county boundaries over time and the fact that many people simply report their place of birth as "London" in the 1911 census.

2. Drop all pairs of linked observations that do not have matching Phonex codes or county of birth, while allowing discrepancies in the reported age of up to 5 years.

3. Compute the Jaro-Winkler score between the first names and last names in each pair of observations. Discard all pairs with a Jaro-Winkler score less than 0.75 for either the first or last name.[31]

4. For each record in the earlier census, determine the maximum Jaro-Winkler score averaged over the first and last names, and the minimum discrepancy in age among all records identified in Step 2. Count the number of records in the later census with a Jaro-Winkler score (J_s) satisfying $(1+0.1)\,J_s \geqslant \bar{J}$, where $\bar{J}$ is the Jaro-Winkler score of the best match, and having a reported age within one year of the closest match.

5. Prioritize linked observations that match on birth parish.

6. Drop all pairs of linked observations with a discrepancy in reported age greater than the minimum discrepancy across all later-year census records matched to an earlier-year census record.

7. Drop any remaining pairs of linked records with a Jaro-Winkler score (J_s) satisfying $(1+0.1)\,J_s < \bar{J}$, where $\bar{J}$ is the Jaro-Winkler score of the best match. In other words, we consider a record uniquely matched on name-age combinations if it is sufficiently "better" than the next closest match.

[30] While the linking methods used in this paper are not exact replications of traditional methods, the approach of incorporating features from different methods is validated by the findings of Bailey et al. (2017) that using a combination of samples generated with the Ferrie (1996) and Feigenbaum (2016) methods results in a much lower Type I error rate.

[31] Economic historians have preferred the Jaro-Winkler score as a string distance measure for linking names across censuses because it places greater weight on characters that match at the beginning of a string (Feigenbaum 2016; Mill and Stein 2016). Jaro-Winkler scores range from 0 to 1, where a score of 0 indicates no common letters, while a score of 1 indicates a perfect match. For more details on the Jaro-Winkler method, and other string comparison algorithms, see Christen (2012).

8. Keep all pairs of linked records with a Jaro-Winkler score greater than 0.80 averaged across the first and last name, that satisfy the following conditions: each earlier-year census has a unique match in the later-year census, and each later-year census record has a unique match in the earlier-year. We exclude records that have unique name-age combinations if the second best match is sufficiently similar.

We present linkage rates separately for all census-to-census linkages in panel A of Table A8. The share of unique matches ranges from 49 to 58 percent across the set of census pairs, with higher match rates for censuses that are closer together, especially those that are only 10 years apart. The census-to-census linkage rates typically found in the literature using complete-count US census data are somewhat lower. This difference can be explained in part by applications with longer windows of time between censuses, typically 30 to 40 years, where sample attrition is of greater concern. Furthermore, the U.S. censuses have less precise information on birth place, at the state level instead of county or parish, which reduces the likelihood of finding unique matches.

A.2 Hospital-to-census linkage

The procedure for linking inpatient hospital admission records to population censuses follows the steps outlined above for census-to-census linkages with a few important modifications. First, we do not observe place of birth in the hospital records and thus do not use it as a linking variable. Second, we do not require that each census record is linked to a unique admission record, given that we do not observe a patient identifier and some patients may be admitted multiple times. Instead, we treat multiple admission records that match to the same census records as belonging to the same person.

We link each hospital admission record to the 1881, 1891 and 1901 censuses provided that the admission occurred within 10 years of the census enumeration date. We use information on the age in years on the day of the hospital admission to determine the age in years on the days of census enumeration. We require the age to differ by no more than 3 years between sources, which is a tighter window than the census-to-census match since age heaping is less of a concern when individuals are observed as children in both the census and hospital records. In the absence of information on place of birth, we prioritize linkages of records that match on county of residence, but we do not require either district or county of residence to match since individuals moved often, even in short time windows between hospital admission and census enumeration.

We discuss the overall linkage rates from hospital records to censuses in Section 3.2. In Table A16 we present the share of hospital admissions that remain after each stage in the matching

procedure, separately by hospital and census year. Overall, the linkage rates are very similar across the three hospitals at each stage of the procedure.

A.3 Multiple linkages to a sample of unique individuals

In order to execute our empirical strategy we must perform three separate linkages:

1. Patients, from hospital admission records to childhood census records
2. Patients, from childhood census record to census record during adulthood
3. Siblings, from childhood census record to census record during adulthood

As a substantial portion of the starting sample is lost through multiple linkages, we must compensate by pooling together multiple hospital-to-census and census-to-census linkages. This section describes the procedure used to identify which records belong to the same individual, and which linked records to use in the analysis for a given individual.

As described in Section 3.1, the hospital admission records do not include a unique patient identifier. We start by assuming that separate admissions belong to the same person if the surname, first name, middle name, implied birth year, and registration district of residence all match across a set of admission records. We use the grouping of records based on these variables as a proxy patient identifier.

Among those patients linked to census records during childhood, we update the unique identifiers based on the census linkages. In a small number of cases, admissions of patients with different proxy identifiers are linked to the same census record in either 1881, 1891, or 1901, and we consider them to be the same individual. When we conduct the second linkage to census records during adulthood, we further consolidate the proxy identifiers. For example, if one admission record is linked to the 1881 census, and another record is linked to the 1891 census, and both census records are linked to the same individual in either the 1901 or 1911 census, then we consider the two admission records to belong to the same patient. As illustrated in Table 1, many patients are linked to more than one census, with hospital-to-census linkage rates ranging from 25 to 28 percent for each of the 1881 through 1901 censuses, but only 38 percent of patients matched to any census.

To select the patient and census record pair to use in the analysis of long-run occupational outcomes, we implement an algorithm which prioritizes linkages according to the following criteria:

1. Choose the census closest to the admission year.
2. Select the census record with the smallest deviation in age between the hospital admission record and the childhood census.
3. Choose the childhood census record linked to the latest census year during adulthood (1901 or 1911).
4. Choose the earliest childhood census record (1881, 1891 or 1901).

Upon completion of these steps, we update the proxy patient identifiers and repeat the procedure once more. The sample used in the main analysis is formed by pooling together individuals from the three childhood census years (1881, 1891 or 1901) who were linked to either of the adulthood census years (1901 or 1911).

The algorithm for prioritizing a pair of records within a set of census linkages for a given patient when considering school attendance as an outcome differs slightly in comparison to the case of long-run outcomes and proceeds as follows:

1. Choose the census closest to the admission year.
2. Select the census record with the smallest deviation in age between the hospital admission record and the childhood census.
3. Prioritize matches to census records of individuals aged 5 to 12 at enumeration.
4. Choose the most recent census data (1881 or 1891).

This procedure ensures that we choose the highest quality match for the analysis sample, before we impose additional restrictions so that we observe the individual in the census during the compulsory schooling years and after the hospital admission.

A.4 Patient-sibling comparisons

When linking the male siblings of male hospital patients across census years, we attempt to match all siblings within 8 years of age of the patient. In the main regression analysis, we impose some restrictions to limit the sample to comparisons of one patient and one sibling per household:

1. Drop households with multiple patients.

2. Among successfully matched male siblings, keep the sibling who is closest in age to the hospital patient. In the cases where we link both an older and younger sibling with the same age gap in comparison to the hospital patient, we randomly choose one of the two siblings, in order to avoid biasing the sibling fixed effects comparisons to either younger or older siblings.

We test robustness of the results to these additional restrictions in Section 6.

B Health Deficiency Index

In this section we describe the procedure used to construct the health deficiency index introduced in Section 3.1. We start with the set of all admissions of male and female patients from the 1870 to 1902 birth cohorts who were admitted to a hospital between 1870 and 1902. Note that the estimation sample includes patients from the 1891 to 1902 birth cohorts who are excluded from the main analysis since they are too young to have occupational outcomes in the available census years. We clean the cause-of-admission text strings and categorize the information into one of seven groups:

1. Disease or medical condition
2. Symptoms
3. Conditions requiring surgery
4. External factors (e.g. poisoning or collisions)
5. Foreign objects
6. Descriptors of severity
7. Body part

If an individual's admission record reports one or more diseases or medical conditions we take the set of these diagnoses as the cause of admission. If not, we go sequential down the list, adding information until we have assigned a primary diagnosis to all possible individuals.

For each diagnosis, we compute its frequency and observed inpatient mortality rate by gender. Then, for individuals with multiple diagnoses, we choose the diagnosis with the highest mortality rate. We break ties by choosing the most frequently occurring diagnosis. This procedure leaves us with a single primary diagnosis per admission record.

Next, we estimate the following linear probability model separately by gender and save the residuals:

$$\mathrm{P}(\textit{Death in hospital})^{g}_{nhay} = \alpha + \theta_h + \delta_a + \gamma_y + \epsilon^{g}_{nhay} \tag{3}$$

where n^g indexes individual in-patient admissions for gender g, h indexes hospitals, a indexes age in years at admission, and y indexes the year of admission. The dependent variable is an indicator that takes the value of one when a patient dies in the hospital. We include hospital (θ),

age at admission (δ), and year of admission (γ) fixed effects. We save the residuals $\widehat{\epsilon_n^g}$ from the regression to use as an input in the next step of computing the health deficiency index.

The estimation excludes observations with no diagnosis and diagnoses with at least 25 observations for which there is no variation in observed inpatient mortality.[32] Next, we assign patients the average residual mortality risk for their primary diagnosis as a proxy for childhood health. For each diagnosis d_j^g of gender g, we compute the following:

$$H_j^g = \frac{\sum_{n=1}^{N^g} \left(I(d_j^g \in C_n^g) \cdot \widehat{\epsilon_n^g} \right)}{\sum_{n=1}^{N^g} \left(I(d_j^g \in C_n^g) \right)}$$

which is the average unexplained mortality risk across all admission of gender g containing diagnosis d_j^g. Finally, we compute the health deficiency index by the following steps:

1. Among diagnoses for which the average residual mortality H_j^g was computed, we construct a max-min standardized score according to:

$$Z_j^g = \frac{H_j^g - \min(H_j^g)}{\max(H_j^g) - \min(H_j^g)}$$

2. For diagnoses with at least 25 observations by gender and no observed variation in inpatient mortality, we assign $Z_j^g = 1$ if all patients died in the hospital and $Z_j^g = 0$ if no patients died in the hospital.

[32] We take 25 observations as the threshold at which we are confident that the cause of admission is certain not to result in a death in the hospital. There are no causes of admission with more than 25 observations for which all patients die in the hospital. Results are similar when we use a threshold of 10 or 50 observations.

C Appendix Tables

Table A1: Common causes of admission in hospital population and the final sample

	Hospital population				Hospital male population				Final sample	
Cause of admission	Frequency	Percent	Mortality rate	Cause of admission	Frequency	Percent	Mortality rate	Cause of admission	Frequency	Percent
Abscess	3,202	4.45	0.04	Abscess	1,901	4.52	0.04	Abscess	112	4.83
Diphtheria	2,773	3.85	0.38	Pneumonia	1,513	3.59	0.11	Fracture	72	3.10
Tubercular Disease	2,396	3.33	0.04	Diphtheria	1,499	3.56	0.36	Pneumonia	71	3.06
Pneumonia	2,368	3.29	0.11	Tubercular Disease	1,422	3.38	0.04	Bronchitis	58	2.50
Chorea	2,104	2.92	0.01	Bronchopneumonia	1,120	2.66	0.32	Phimosis	58	2.50
Bronchopneumonia	1,937	2.69	0.30	Bronchitis	999	2.37	0.17	Typhoid Fever	48	2.07
Bronchitis	1,766	2.45	0.16	Fracture	939	2.23	0.03	Chorea	47	2.03
Fracture	1,251	1.74	0.02	Meningitis	718	1.71	0.77	Diphtheria	45	1.94
Meningitis	1,226	1.70	0.77	Empyema	686	1.63	0.14	Tubercular Disease	44	1.90
Cleft Palate	1,093	1.52	0.00	Chorea	618	1.47	0.01	Injury	41	1.77
Empyema	1,090	1.51	0.13	Fever	613	1.46	0.11	Empyema	40	1.72
Typhoid Fever	1,006	1.40	0.06	Phimosis	601	1.43	0.01	Rheumatism	38	1.64
Fever	1,003	1.39	0.11	Injury	581	1.38	0.04	Cleft Palate	38	1.64
Tuberculosis	983	1.37	0.57	Typhoid Fever	569	1.35	0.07	Talipes	35	1.51
Morbus Cordis	954	1.33	0.17	Harelip	556	1.32	0.02	Fever	33	1.42
Rheumatism	931	1.29	0.02	Tuberculosis	539	1.28	0.57	Necrosis	33	1.42
Harelip	852	1.18	0.02	Cleft Palate	535	1.27	0.01	Harelip	33	1.42
Talipes	799	1.11	0.01	Rheumatism	524	1.24	0.02	Rickets	32	1.38
Rickets	794	1.10	0.04	Talipes	517	1.23	0.00	Pleurisy	31	1.34
Injury	776	1.08	0.03	Morbus Cordis	456	1.08	0.16	Morbus Cordis	28	1.21
Phthisis	775	1.08	0.23	Burn	451	1.07	0.25	Disease Knee	25	1.08
Burn	773	1.07	0.27	Rickets	440	1.05	0.05	Diarrhea	25	1.08
Diarrhea	713	0.99	0.26	Diarrhea	418	0.99	0.24	Scarlet Fever	25	1.08
Disease Hip	704	0.98	0.02	Laryngitis	412	0.98	0.16	Eczema	24	1.03
Necrosis	679	0.94	0.03	Necrosis	409	0.97	0.03	Disease Hip	24	1.03
Total (top 25)	32,948	45.76	0.16	Total (top 25)	19,036	45.22	0.15	Total (top 25)	1,060	45.70
Outside top 25	39,006	54.24	0.11	Outside top 25	23,064	54.78	0.11	Outside top 25	1,260	54.30

Notes: This table lists the 25 most common causes of admission in the hospital population, the population of hospitalized males, and the final sample used in the analysis. The hospital population consists of all admissions by male and female patients born between 1870 and 1902 and admitted at ages 0 to 11 between 1870 and 1902 at GOSH, Barts, or Guy's Hospitals. The causes of admissions are tabulated after cleaning the text strings transcribed from the admissions registers. The mortality rate refers to the share of admissions in which a patient died in the hospital. The final sample refers to the set of hospital admissions corresponding to the 1998 male patients included in the main analysis in columns 1 to 3 of Table 2. The mortality rate is not shown for the final sample since it only includes patients who survived until adulthood.

45

Table A2: Common occupational titles by occupational class

	(1) White collar	(2) Skilled	(3) Semi-skilled	(4) Unskilled
1	Clerk	Carpenter	Carman	General Labourer
2	Railway Clerk	Cabinet Maker	Coal Miner Hewer	Labourer
3	Commercial Clerk	Bricklayer	House Painter	Farm Labourer
4	Police Constable	French Polisher	Postman	Gardener Domestic
5	Insurance Agent	Butcher	Porter	Railway Porter

Notes: This table lists the five most common occupations in each of four occupational classes for the final sample of patients and siblings used in the main regressions. Column (1) combines professional, managerial and clerical occupations (Classes 1 and 2 in the seven category HISCLASS scheme) into a white collar class, Column (2) subsumes farmers into skilled workers (HISCLASS 3 and 4), Column (3) displays low-skilled workers (HISCLASS 5), and Column (4) combines unskilled workers as well as low and unskilled farm workers (HISCLASS 6 and 7).

Table A3: Number of beds and admissions to hospitals in London, 1893

Hospital in 1894	# Beds	Inpatients	Outpatients	Inpatient %
		Panel A: General hospitals		
Barts	675	6,474	159,802	4.05
Guy's	695	6,325	57,223	11.05
Top-12 General	4,937	52,231	688,187	7.59
Barts share (%)	13.7	12.4	23.2	
Guy's share (%)	14.1	12.1	8.3	
		Panel B: Children's hospitals		
GOSH	178	1,801	27,334	6.59
Top-6 Children's	497	6,281	110,386	5.69
GOSH share (%)	35.8	28.7	24.8	

Notes: This table displays the number of hospital beds, the number of inpatients, the number of outpatients, and the share of inpatients among outpatients from 1894 for hospitals used in the analysis. The original source does not indicate whether inpatients are included in the outpatient totals. The table also shows the shares for the sample hospitals relative to the twelve largest general and children's hospitals in London.
Source: Chatto and Windus (1897).

Table A4: Descriptive statistics for hospital catchment areas (1891 census)

	(1) Barts	(2) GOSH	(3) Guys	(4) Rest of London
Share aged 0 to 4	0.121	0.110	0.131	0.119
Share aged 5 to 11	0.149	0.136	0.158	0.149
Sibship size	4.029	3.906	4.081	4.108
Share age 0 to 11 living with mother	0.920	0.913	0.925	0.912
Share age 0 to 11 living with father	0.864	0.853	0.875	0.857
Share of unskilled fathers	0.109	0.120	0.202	0.157
Share of unskilled household heads	0.103	0.110	0.193	0.148
Share of household heads married	0.865	0.844	0.884	0.869
Share of immigrants	0.083	0.103	0.070	0.102
Catchment area size (N)	589,024	1,174,261	341,354	2,722,614

Notes: This table presents descriptive statistics from the 1891 Census of England for the catchment areas of each hospital used in the analysis. A hospital's catchment area is defined as the set of registration districts from which the most patients are admitted and which together account for at least 50 percent of total admissions by children age 0 to 11. The Barts Hospital catchment area includes: Holborn, Shoreditch, and Islington. GOSH catchment area includes: Holborn, Islington, Pancras, Kensington, Marylebone, Shoreditch, and St Giles. The Guy's Hospital catchment area includes: St Olave Southwark and St Saviour Southwark. Results are similar when using the 1881 or 1901 census.

Table A5: Selection into hospitalization for male patients

	Observed in hospital records [$\times$ 100]		
	(1)	(2)	(3)
Father skilled	0.061***	0.039	0.038
	(0.020)	(0.031)	(0.024)
Father semi-skilled	0.069***	0.053*	0.053**
	(0.020)	(0.027)	(0.022)
Father unskilled	0.089***	0.074*	0.071**
	(0.025)	(0.037)	(0.030)
Mean of Y	0.430	0.430	0.430
Catchment controls	Yes	No	No
District FE	No	Yes	No
Parish FE	No	No	Yes
N	715,103	715,103	715,103

Notes: The sample consists of individuals who were ages 0 to 5 and residing in the County of London when enumerated in the 1881, 1891 or 1901 censuses. The dependent variable is an indicator for a unique match to an inpatient hospital admission that occurred up to 10 years after the census enumeration date and when the individual was age 0 to 11 at the time of admission. The regressions also include age at enumeration by census year fixed effects for patients and their fathers. The sample is further restricted to male children with fathers age 21 to 50 at enumeration.

Point estimates marked ***, **, and * are statistically significant at the 1, 5, and 10 percent levels, respectively.

Table A6: Sibling-specific determimants of hospitalization

	Hospitalization (Patients vs. siblings)				Health deficiency index | Hospitalization (Patients only)			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	1881	1891	1901	Any	1881	1891	1901	Any
First born	−0.073*	−0.158***	−0.054	−0.109***	−0.004	−0.002	−0.001	−0.003
	(0.039)	(0.039)	(0.101)	(0.027)	(0.008)	(0.007)	(0.015)	(0.005)
Female	−0.043**	−0.063***	−0.101**	−0.057***	−0.000	−0.001	0.001	−0.001
	(0.020)	(0.020)	(0.045)	(0.013)	(0.006)	(0.006)	(0.014)	(0.004)
First born × female	0.123**	0.122**	0.007	0.110***	−0.006	0.012	−0.002	0.002
	(0.056)	(0.053)	(0.133)	(0.037)	(0.012)	(0.011)	(0.026)	(0.008)
Mean of Y	0.395	0.426	0.413	0.411	0.300	0.313	0.333	0.310
N	6,868	6,973	1,512	15,353	2,712	2,970	624	6,306

Notes: Columns 1 to 4 present sibling fixed effects estimates with an indicator for hospitalization as the dependent variable, while columns 5 to 8 show OLS estimates with the health deficiency index as the dependent variable when restricting to patients only. Linkages from the 1881, 1891, and 1901 censuses, respectively, to hospital records up to 10 years after the census enumeration date are shown in columns 1 to 3 (and 5 to 7). The samples consists of all individuals enumerated at ages 0 to 5 in the County of London in households with at least one patient admitted to the hospital at ages 0 to 11 years old no more than 10 years after the census enumeration date. Columns 4 and 8 pool together the samples in the preceding columns. All regressions include age-at-enumeration by census year fixed effects. Columns 1 to 4 cluster standard errors by household while columns 5 to 8 report heteroskedasticity robust standard errors.

Point estimates marked ***, **, and * are statistically significant at the 1, 5, and 10 percent levels, respectively.

Table A7: Patient health deficiency index and likelihood of linkage to census

	Census year linked to hospital records			
	(1)	(2)	(3)	(4)
	1881	1891	1901	Any
Health deficiency index	−0.097	−0.019	−0.090**	−0.059**
	(0.069)	(0.038)	(0.040)	(0.026)
Mean of Y	0.265	0.272	0.251	0.263
N	4,758	12,355	9,774	26,887

Notes: Columns 1 to 3 present results for linkages from the hospital records to the 1881, 1891, and 1901 censuses, respectively. The samples consists of all patients admitted to the hospital at ages 0 to 11 years old no more than 10 years prior to the census enumeration date and discharged from the hospital prior to the census enumeration date. Column 4 pools together the samples in the preceding columns.

Point estimates marked ***, **, and * are statistically significant at the 1, 5, and 10 percent levels, respectively.

Table A8: Census-to-census linkage rates and patient vs. sibling differences

	Outcome year = 1901		Outcome year = 1911		
	(1)	(2)	(3)	(4)	(5)
	1881	1891	1881	1891	1901
	Panel A: Census-to-census linkage rates for County of London				
No match	0.332	0.284	0.364	0.319	0.258
Multiple matches	0.169	0.180	0.151	0.170	0.164
Unique match	0.498	0.535	0.485	0.511	0.579
Baseline sample	814,157	976,877	814,157	976,877	1,059,496
	Panel B: Effects of hospital admission on match success				
Patient	0.023	0.024**	0.010	0.024**	0.020
	(0.014)	(0.010)	(0.015)	(0.011)	(0.015)
% effect	4.4	4.1	2.0	4.3	3.0
Mean of Y	0.531	0.597	0.508	0.559	0.653
N	5,447	8,338	5,447	8,338	4,432

Notes: Panel A presents census-to-census linkage rates for boys residing in the County of London in the base-year census. See appendix A.1 for a description of the linkage procedure. Panel B presents sibling fixed effects estimates with indicators for a unique match between censuses as the dependent variable. The sample includes male patients linked from the hospital records to a census during childhood and their brother closest in age. Each regression includes age-at-enumeration and birth order fixed effects, as well as standardized measures of first name frequency and an interaction of first name and surname frequency. Standard errors are clustered by childhood household.

Point estimates marked ***, **, and * are statistically significant at the 1, 5, and 10 percent levels, respectively.

Table A9: Scaling of health estimates by intergenerational transmission of status

	(1) White collar	(2) Skilled +	(3) Semi-skilled +
	Panel A: Effect of σ change in health deficiency index		
Health deficiency index	−0.016***	−0.015**	−0.013***
	(0.006)	(0.007)	(0.005)
% effect	5.8	2.8	1.6
Mean of Y	0.273	0.544	0.850
N	3,996	3,996	3,996
	Panel B: Intergenerational occupational elasticities		
Father's status	0.251***	0.220***	0.184***
	(0.001)	(0.001)	(0.001)
Mean of Y	0.240	0.488	0.805
N	2,004,077	2,004,077	2,004,077
	Panel C: Scaled effects (%)		
Health deficiency index (σ)	6.3	7.0	7.3
Patient	16.2	18.1	18.6

Notes: Estimates in Panel A represent the effect of a 1 s.d. change in the health deficiency index. Panel B presents estimates using data on males aged 0 to 11 linked from the 1881 to the 1911 complete-count census. In each column, *Father's status* is an indicator for the father's occupational status equivalent to the dependent variable. The regressions also control for an indicator for above-median sibship size, match quality dummies, as well as own and father's birth year fixed effects. The sample restricts attention to fathers and sons no more than 20 years apart in age when occupational outcomes are observed. Panel C displays the coefficients in Panel A and the coefficients on the hospitalization indicator from columns 1 to 3 in Table 2 scaled by the estimates in Panel B.

Point estimates marked ***, **, and * are statistically significant at the 1, 5, and 10 percent levels, respectively.

Table A10: Intergenerational mobility matrix, sample vs. population

	Father's occupational class							Total	N
	Professional	Clerical	Skilled	Farmer	Semi-skilled	Unskilled	Farm laborer		
				Panel A: Sons in 1901/1911 (Patients and siblings, final sample)					
Professional	0.0	5.5	3.4	5.6	3.7	1.4	0.0	3.6	95
Clerical	0.0	41.4	15.9	22.2	23.3	17.7	0.0	23.4	611
Skilled	0.0	18.9	40.3	16.7	20.2	24.7	0.0	26.9	701
Farmer	0.0	1.1	0.6	20.4	1.8	0.3	0.0	1.5	38
Semi-skilled	0.0	23.1	29.3	18.5	38.3	34.6	0.0	31.7	827
Unskilled	0.0	7.6	9.1	1.9	11.1	19.7	0.0	10.8	282
Farm laborer	0.0	2.5	1.4	14.8	1.6	1.7	0.0	2.0	52
Total	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	2,606
N	0	476	832	54	888	356	0	2,606	
				Panel B: Sons age 18 to 41 in 1911					
Professional	20.5	7.8	3.9	4.1	3.4	2.8	2.0	4.6	75,182
Clerical	30.0	37.9	18.6	16.4	15.9	15.2	11.7	19.4	319,758
Skilled	15.7	18.3	35.2	12.1	18.3	18.9	13.2	21.2	350,483
Farmer	2.5	2.1	1.8	32.6	1.6	2.2	5.6	3.9	64,141
Semi-skilled	22.3	23.4	28.3	18.5	47.8	35.8	29.4	34.0	560,377
Unskilled	6.3	7.8	9.1	5.1	9.6	18.0	11.4	9.9	163,764
Farm laborer	2.8	2.6	3.2	11.3	3.3	7.1	26.7	7.0	116,279
Total	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	1,649,984
N	64,047	218,500	386,781	83,751	521,179	156,829	218,897	1,649,984	

Notes: This table displays the transition matrix for intergenerational mobility between fathers and sons in the final sample and the population. Panel A shows the occupational transition probabilities for patients and siblings in the final sample observed in the 1901 or 1911 census, in comparison to their fathers. Panel B presents occupational transition probabilities for fathers and sons across England from the 1881 to 1911 linked sample using the complete count files. The linked sample restricts attention to sons age 18 to 41 in 1911 and fathers observed in 1881 within 20 years of the son's age in 1911.

Table A11: Benchmark for effects on intergenerational mobility

	(1)	(2)	(3)	(4)
	Linked sample		England, 1881-1911	
	Class ↘	Class ↗	Class ↘	Class ↗
	Panel A: Estimates of mobility rates			
Sample mean	0.279	0.721	0.267	0.733
	Panel B: Effects of poor health scaled by mobility rates			
Patient	0.139	0.055	0.146	0.054
Health deficiency index (σ)	0.050	0.019	0.052	0.019

Notes: Columns 1 and 2 of panel A present means of the dependent variable in the estimation sample from columns 4 and 5 in Table 2. The dependent variables are indicators for an individual who enters a lower occupational class than his father (column 1) or an individual with the same or higher occupational class (column 2). Columns 3 and 4 show the sample means of downward and of upward mobility from the population of fathers linked from the 1881 census to sons age 18 to 41 in the 1911 censuses. Columns 1 and 2 of panel B presents coefficients from columns 4 and 5 of Table 2 scaled by the sample means in Panel A.

Table A12: Hetergeneity by severity, age at admission, and baseline SES

	(1) Scholar	(2) White collar	(3) Skilled +	(4) Semi-skilled +	(5) Class ↘	(6) Class ↗
	Panel A: Interaction with above vs. below median HDI					
Patient × low-HDI	−0.023***	−0.045***	−0.032*	−0.030**	0.047**	−0.049**
	(0.007)	(0.017)	(0.019)	(0.014)	(0.020)	(0.020)
Patient × high-HDI	−0.015*	−0.035*	−0.049**	−0.039**	0.029	−0.027
	(0.009)	(0.019)	(0.023)	(0.016)	(0.022)	(0.022)
P-value	0.415	0.687	0.552	0.666	0.518	0.440
	Panel B: Interaction with early (0-4) vs. late (5-11) childhood admissions					
Patient × [0-4]	−0.023***	−0.053**	−0.057**	−0.039**	0.036	−0.038
	(0.008)	(0.021)	(0.024)	(0.017)	(0.024)	(0.024)
Patient × [5-11]	−0.015**	−0.033**	−0.028	−0.032**	0.037*	−0.034*
	(0.007)	(0.017)	(0.019)	(0.013)	(0.019)	(0.019)
P-value	0.370	0.451	0.334	0.732	0.978	0.891
	Panel C: Interaction with high vs. low parental SES					
Patient × low-SES	−0.033***	−0.049***	−0.030	−0.018	0.033*	−0.033*
	(0.008)	(0.019)	(0.022)	(0.017)	(0.017)	(0.017)
Patient × high-SES	−0.013*	−0.032*	−0.037*	−0.040***	0.045*	−0.045*
	(0.007)	(0.019)	(0.021)	(0.014)	(0.024)	(0.024)
P-value	0.035	0.514	0.822	0.285	0.678	0.651
Mean of Y	0.745	0.273	0.544	0.850	0.279	0.721
N	4,832	3,996	3,996	3,996	2,606	2,606

Notes: Panel A presents estimates in which we interact the indicator for hospital patients with indicators for being admitted for conditions with above and below median health deficiency index. Panel B interacts the indicator variable for hospitalization with separate indicators for early- (age 0 to 4) and late-childhood (age 5 to 11) admission, which are coded based on a patient's first observed admission to a hospital. Panel C presents results from interacting the hospital indicator with indicators for high and low parental SES, where high SES is defined as having a father with a white collar or skilled occupation. In column 1, the dependent variable is an indicator for school attendance in the sample of mixed gender patient vs. sibling comparisons corresponding to column 5 of Table 3. Columns 2 to 6 present estimates for the outcomes variables shown in Table 2. Standard errors are clustered by childhood household.

Point estimates marked ***, **, and * are statistically significant at the 1, 5, and 10 percent levels, respectively.

Table A13: Long-run outcomes: Robustness to selective mortality

	(1) Baseline Estimate	(2) Drop high mortality	(3) Drop low mortality	(4) Drop infant admission	(5) Drop multiple admissions	(6) Drop contagious
			Panel A: Effects on P(White collar)			
Health deficiency index	−0.122***	−0.159***	−0.125***	−0.135***	−0.160***	−0.164***
	(0.044)	(0.051)	(0.045)	(0.047)	(0.048)	(0.053)
Mean of Y	0.273	0.270	0.279	0.276	0.268	0.275
			Panel B: Effects on P(Skilled +)			
Health deficiency index	−0.120**	−0.162***	−0.123**	−0.097*	−0.124**	−0.141**
	(0.053)	(0.058)	(0.054)	(0.054)	(0.056)	(0.061)
Mean of Y	0.544	0.545	0.546	0.550	0.539	0.542
			Panel C: Effects on P(Semi-skilled +)			
Health deficiency index	−0.104***	−0.124***	−0.102***	−0.112***	−0.105***	−0.141***
	(0.037)	(0.041)	(0.037)	(0.038)	(0.039)	(0.044)
Mean of Y	0.850	0.849	0.853	0.851	0.852	0.850
N	3,996	3,638	3,594	3,628	3,528	3,282
			Panel D: Effects on P(Class ↘)			
Health deficiency index	0.108**	0.102*	0.118**	0.092*	0.123**	0.092
	(0.053)	(0.058)	(0.054)	(0.055)	(0.055)	(0.062)
Mean of Y	0.279	0.275	0.283	0.278	0.279	0.283
			Panel E: Effects on P(Class ↗)			
Health deficiency index	−0.106**	−0.109*	−0.114**	−0.096*	−0.118**	−0.107*
	(0.053)	(0.058)	(0.054)	(0.055)	(0.055)	(0.062)
Mean of Y	0.721	0.726	0.717	0.721	0.721	0.718
N	2,606	2,360	2,346	2,356	2,310	2,126

Notes: This table is identical to Table 4 with the exception that the treatment variable is changed from an indicator for hospitalization to the continuous health deficiency index. Standard errors are clustered by childhood household. Point estimates marked ***, **, and * are statistically significant at the 1, 5, and 10 percent levels, respectively.

Table A14: Long-run outcomes: Robustness to sample selection

	(1) Baseline Estimate	(2) Add multiple siblings	(3) Add multiple patient hhlds.	(4) County of London only	(5) Drop Guy's Hospital	(6) Unique within census county	(7) Hospital-census county match
				Panel A: Effects on P(White collar)			
Health deficiency index	−0.122***	−0.109***	−0.106***	−0.119**	−0.132***	−0.129***	−0.157***
	(0.044)	(0.042)	(0.041)	(0.048)	(0.048)	(0.047)	(0.056)
Mean of Y	0.273	0.274	0.274	0.281	0.278	0.277	0.287
				Panel B: Effects on P(Skilled +)			
Health deficiency index	−0.120**	−0.094*	−0.085*	−0.117**	−0.150***	−0.111**	−0.159**
	(0.053)	(0.049)	(0.048)	(0.057)	(0.056)	(0.055)	(0.065)
Mean of Y	0.544	0.543	0.543	0.551	0.553	0.544	0.554
				Panel C: Effects on P(Semi-skilled +)			
Health deficiency index	−0.104***	−0.078**	−0.079**	−0.109***	−0.101***	−0.090**	−0.072
	(0.037)	(0.034)	(0.034)	(0.039)	(0.039)	(0.039)	(0.046)
Mean of Y	0.850	0.848	0.849	0.860	0.857	0.850	0.861
N	3,996	4,828	4,870	3,270	3,596	3,578	2,568
				Panel D: Effects on P(Class ↘)			
Health deficiency index	0.108**	0.085*	0.075*	0.095*	0.120**	0.083	0.091
	(0.053)	(0.044)	(0.043)	(0.057)	(0.057)	(0.055)	(0.065)
Mean of Y	0.279	0.280	0.282	0.281	0.278	0.278	0.276
				Panel E: Effects on P(Class ↗)			
Health deficiency index	−0.106**	−0.085*	−0.075*	−0.093	−0.118**	−0.084	−0.098
	(0.053)	(0.044)	(0.043)	(0.057)	(0.057)	(0.056)	(0.064)
Mean of Y	0.721	0.720	0.718	0.720	0.722	0.722	0.724
N	2,606	4,021	4,053	2,198	2,334	2,292	1,700

Notes: This table is identical to Table 5 with the exception that the treatment variable is changed from an indicator for hospitalization to the continuous health deficiency index. Standard errors are clustered by childhood household.

Point estimates marked ***, **, and * are statistically significant at the 1, 5, and 10 percent levels, respectively.

Table A15: Robustness in schooling outcomes

	(1) Column (5) Table 3	(2) Drop high mortality	(3) Drop low mortality	(4) Drop infant admissions	(5) Drop multiple admissions	(6) Drop contagious
			Panel A: Selective mortality and scarring			
Patient	−0.020***	−0.019***	−0.020***	−0.018***	−0.020***	−0.020***
	(0.006)	(0.006)	(0.006)	(0.007)	(0.006)	(0.006)
Health deficiency index	−0.061***	−0.062***	−0.062***	−0.055**	−0.063***	−0.058***
	(0.020)	(0.023)	(0.021)	(0.022)	(0.022)	(0.022)
Mean of Y	0.745	0.756	0.745	0.752	0.753	0.752
N	4,832	4,344	4,360	4,242	4,244	3,970
	(7) Add multiple siblings	(8) Add multiple patient hhlds.	(9) County of London only	(10) Drop Guy's Hospital	(11) Unique within census county	(12) Hospital-census county match
			Panel B: Sample selection and definition of treatment			
Patient	−0.020***	−0.020***	−0.022***	−0.023***	−0.017***	−0.021***
	(0.006)	(0.006)	(0.007)	(0.006)	(0.006)	(0.007)
Health deficiency index	−0.057***	−0.056***	−0.060***	−0.069***	−0.054**	−0.057**
	(0.020)	(0.020)	(0.021)	(0.022)	(0.021)	(0.024)
Mean of Y	0.744	0.744	0.731	0.745	0.751	0.734
N	5,940	6,002	3,874	4,424	4,436	3,150

Notes: Each cell comes from a separate regression. Column 1 of panel A reproduces the main estimates for the schooling outcomes from column 5 of Table 3, which is based on a sample of individuals aged 5 to 10 when enumerated in the census. See Table 3 for a list of variables included in the regressions. See Tables 4 and 5 for a description of the sample restrictions in the remaining columns and panels. Standard errors are clustered by childhood household.

Point estimates marked ***, **, and * are statistically significant at the 1, 5, and 10 percent levels, respectively.

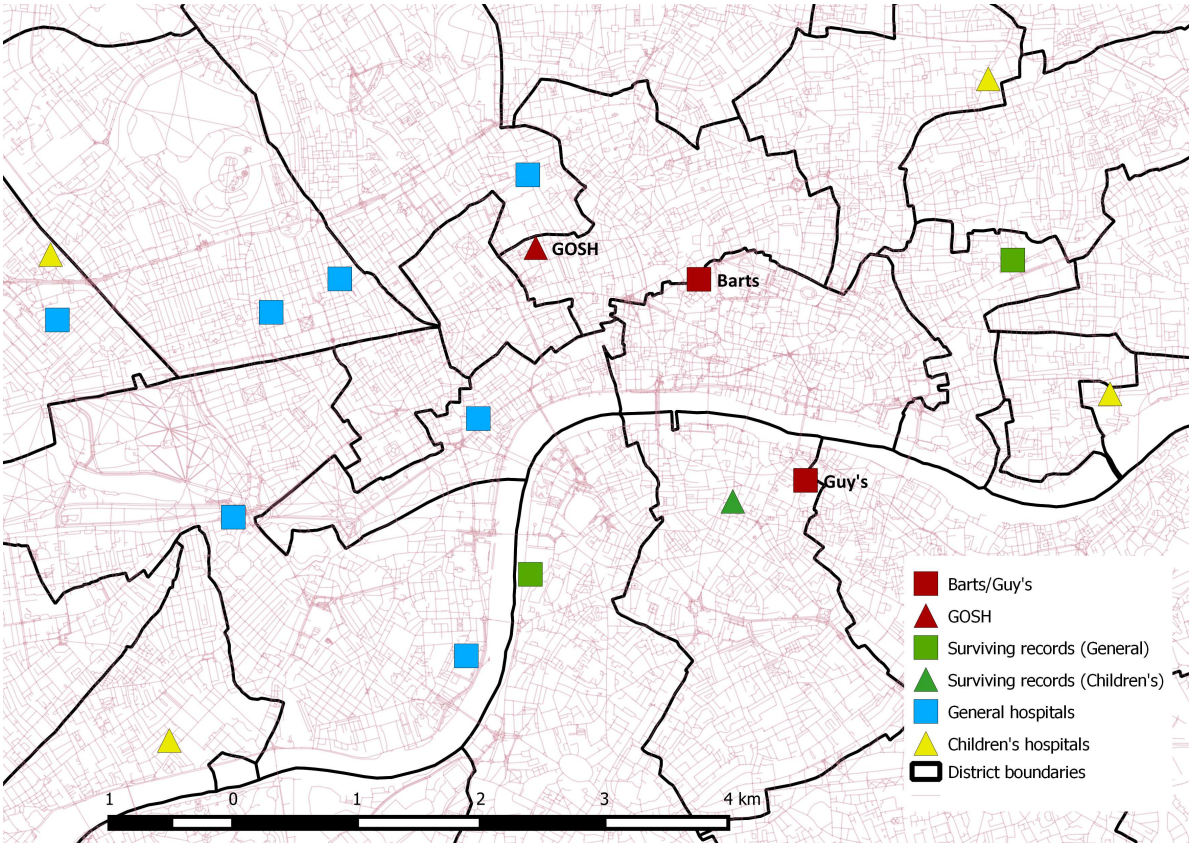
Table A16: Linkage rates to census records by hospital (males)

	Census year linked to hospital records			
	(1) 1881	(2) 1891	(3) 1901	(4) Any
	Panel A: Barts hospital			
No match	0.063	0.050	0.155	0.031
Multiple matches	0.664	0.660	0.582	0.616
Unique match	0.273	0.290	0.263	0.353
Sibling present	0.257	0.265	0.222	0.334
Patient matched	0.192	0.172	0.143	0.232
Patient and sibling	0.078	0.074	0.036	0.117
In final sample	0.063	0.050	0.026	0.086
Total admissions	8,441	11,551	4,850	14,221
	Panel B: GOSH for Sick Children			
No match	0.069	0.064	0.213	0.044
Multiple matches	0.664	0.657	0.559	0.622
Unique match	0.267	0.279	0.229	0.334
Sibling present	0.247	0.251	0.195	0.311
Patient matched	0.181	0.150	0.116	0.204
Patient and sibling	0.073	0.060	0.032	0.099
In final sample	0.059	0.045	0.024	0.078
Total admissions	6,052	9,042	4,524	11,319
	Panel C: Guy's Hospital			
No match	0.045	0.035	0.052	0.018
Multiple matches	0.696	0.692	0.698	0.650
Unique match	0.260	0.273	0.250	0.331
Sibling present	0.243	0.244	0.214	0.309
Patient matched	0.178	0.161	0.125	0.218
Patient and sibling	0.071	0.069	0.036	0.109
In final sample	0.058	0.039	0.029	0.075
Total admissions	1,410	2,530	1,295	2,908

Notes: This table presents linkage rates from the hospital records to the censuses during childhood and adulthood separately by hospital. See Table 1 for a description of each sample restriction.

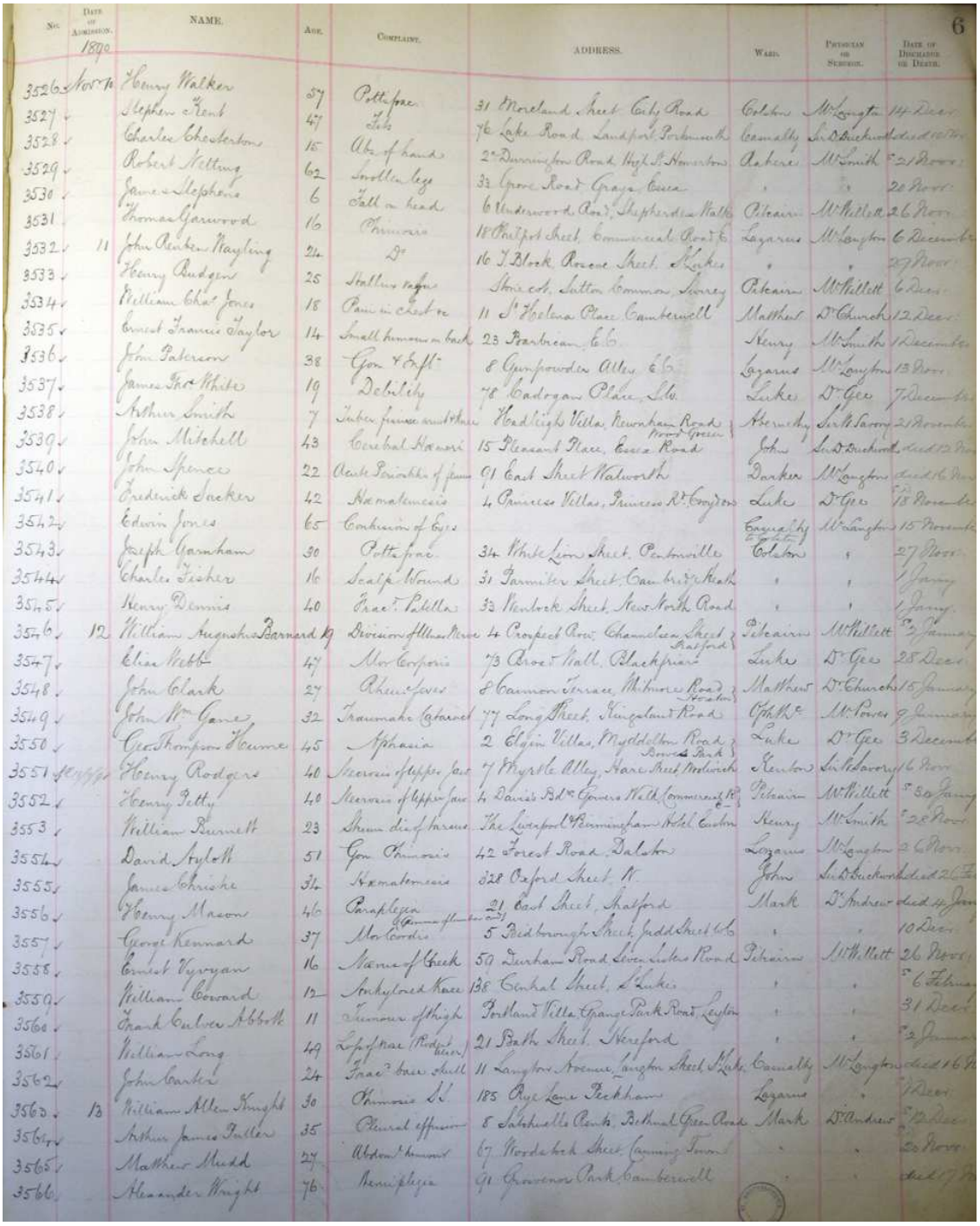
D Appendix Figures

Figure A1: London hospital locations and surviving records



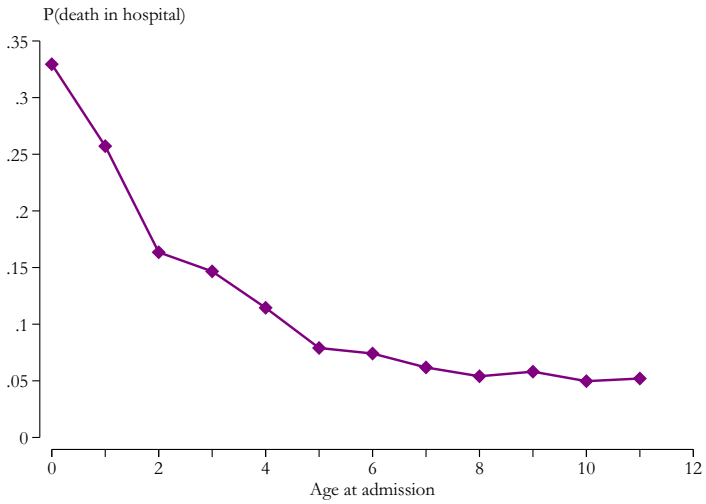
Notes: A map of central London marking the locations of hospitals in empirical sample (red squares and triangle), general hospitals (squares symbol) and children's hospitals (triangle symbol). The subset of these hospitals with surviving archival records is marked in green.

Figure A2: Sample inpatient admission register from St. Bartholomew's Hospital

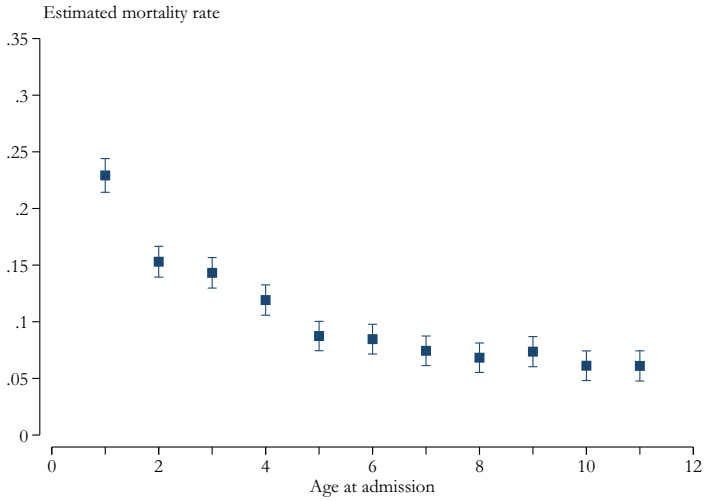


Notes: Sample page from inpatient admission register for St. Bartholomew's Hospital in London. Each page contains the date of admission, the patient's name, age, complaint and address, the name of the ward in which the patient was admitted, the name of the physician or surgeon who treated the patient, and the date of discharge or death. Source: Photographed by authors at St. Bartholomew's Hospital Archive (archival reference number BH/M/3).

Figure A3: In-hospital mortality by age at admission



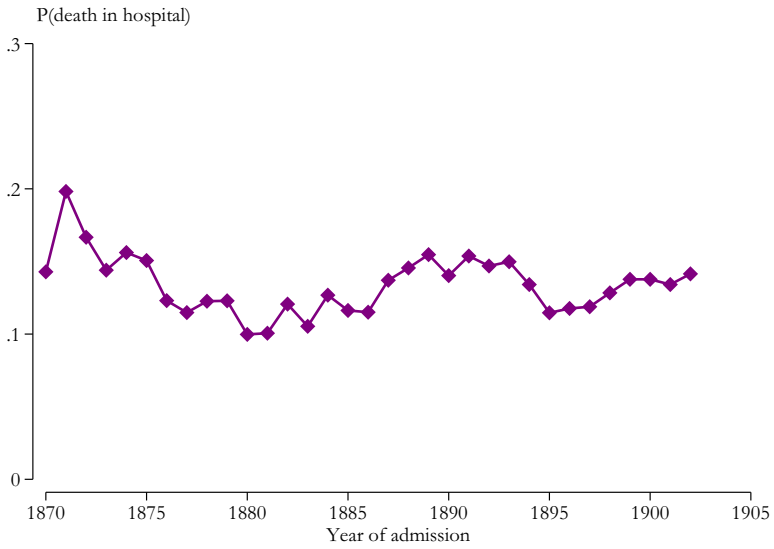
(a) Raw data



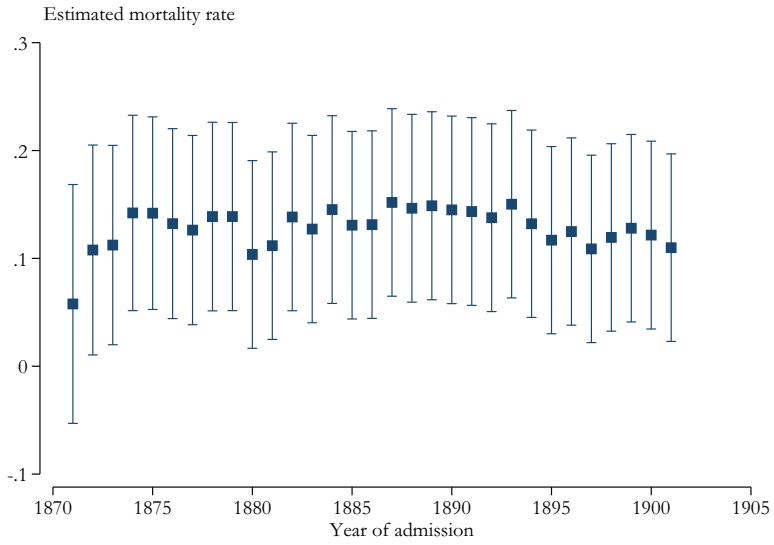
(b) Regression adjusted

Notes: Panel A plots the in-hospital mortality rate by age at admission and Panel B presents regression-adjusted estimates. Panel B plots estimated fixed effects on age at admission (with age 0 as the excluded category) from a linear probability model which also includes admission year, hospital, gender, and number-of-comorbidity fixed effects, as well as indicators for above or below median length of stay, being treated by a doctor, and transferred to another hospital as covariates. We compute predicted mortality for age 0 by setting all age fixed effects and evaluating the remaining variables in the model at their means, then add the predicted value to each of the age fixed effects. The samples include data on all in-patients aged 0 to 11 born between 1869 and 1902, and admitted between 1870 and 1902 to the Hospital for Sick Children at Great Ormond Street, Guy's Hospital, or St. Bartholomew's Hospital, in London.

Figure A4: In-hospital mortality by year of admission



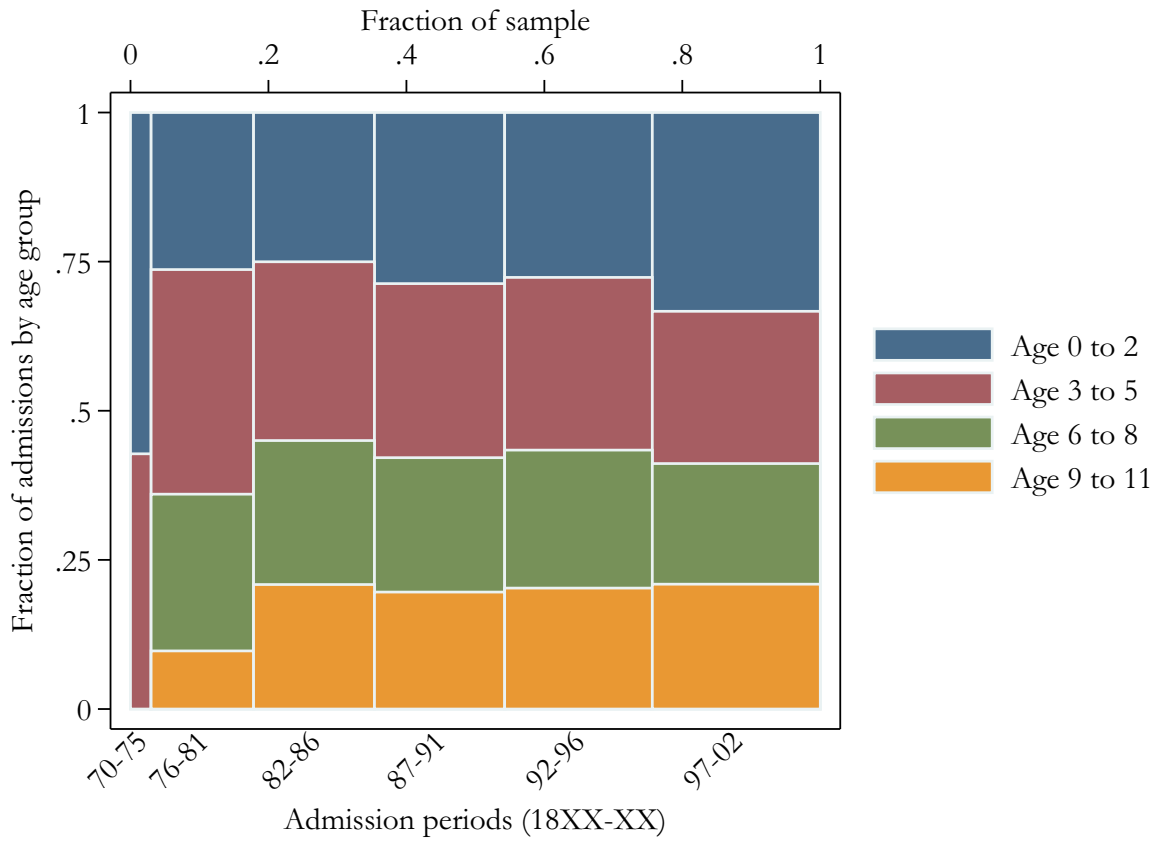
(a) Raw data



(b) Regression adjusted

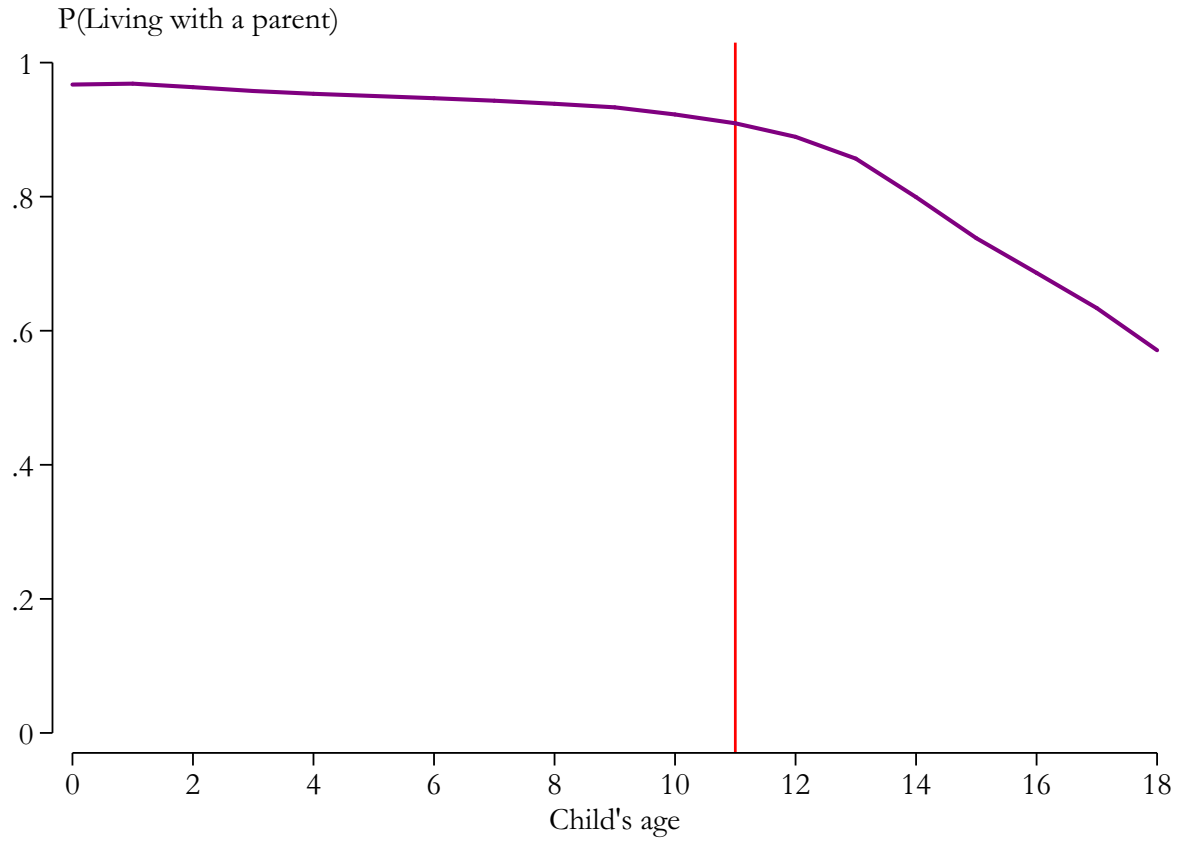
Notes: Panel A plots the in-hospital mortality rate by year of admission and Panel B presents regression-adjusted estimates. Panel B plots estimated fixed effects on year of admission (with 1870 as the excluded category) from a linear probability model which also includes admission age, hospital, gender, and number-of-comorbidity fixed effects, as well as indicators for above or below median length of stay, being treated by a doctor, and transferred to another hospital as covariates. The samples include data on all in-patients aged 0 to 11 born between 1869 and 1902, and admitted between 1870 and 1902 to the Hospital for Sick Children at Great Ormond Street, Guy's Hospital, or St. Bartholomew's Hospital, in London.

Figure A5: Admissions by age group and period



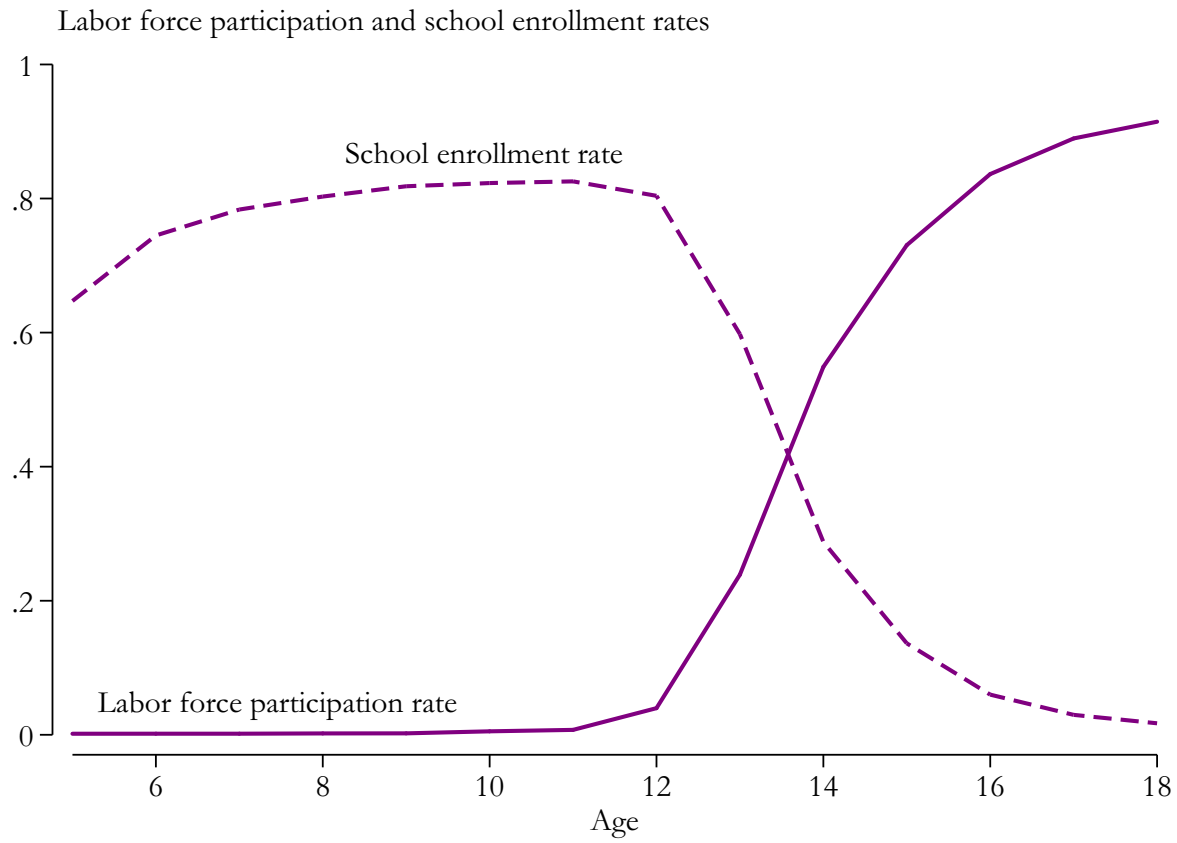
Notes: This figure provides a visualization of in-patient admissions by age group (0 to 2, 3 to 5, 6 to 8, and 9 to 11) and admission period (1870-75, 1876-81, 1882-86, 1887-91, 1892-96, 1897-1902). It uses data on all in-patients aged 0 to 11 admitted between 1870 and 1902 to the Hospital for Sick Children at Great Ormond Street, Guy's Hospital, or St. Bartholomew's Hospital, in London.

Figure A6: Share of children living with a parent in 1881, by age



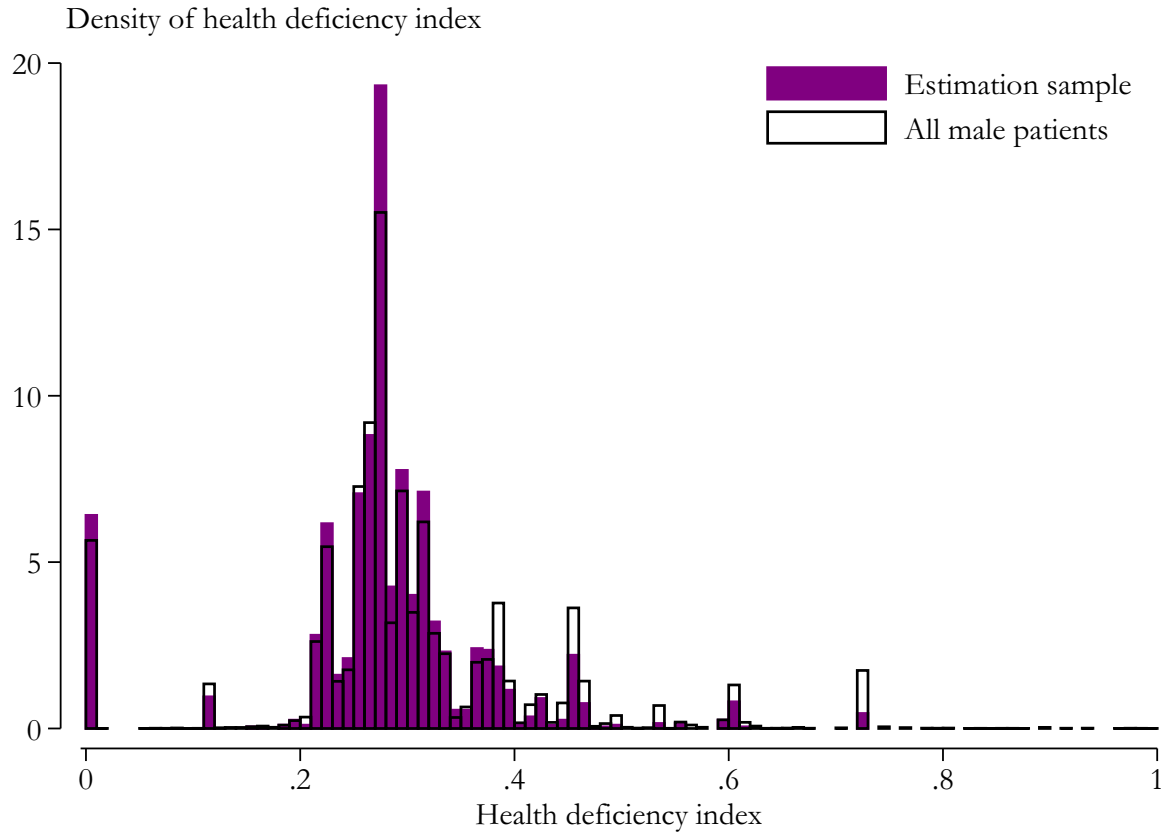
Notes: This figure plots the share of children who were enumerated in the 1881 census in a household in which at least one parent was present. The sample consists of all households in the County of London. Results are similar in the 1891 census.

Figure A7: Labor force participation and school enrollment by age in 1881



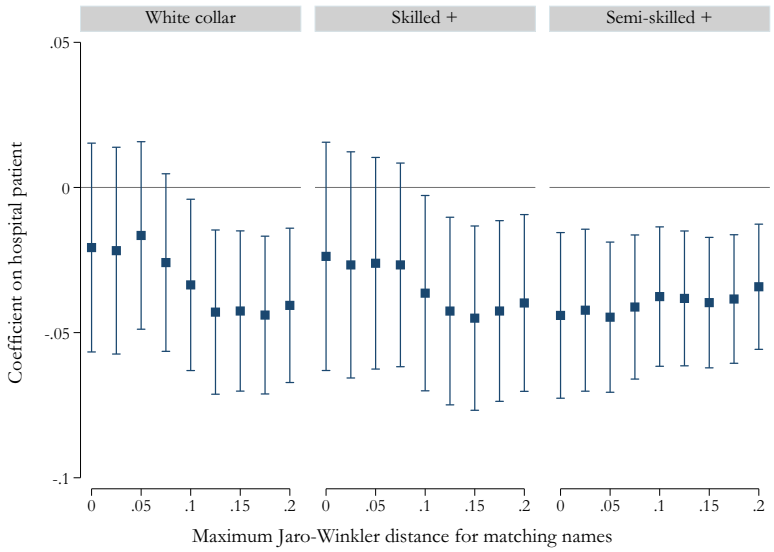
Notes: This figure plots the labor force participation rate (solid lines) and school enrollment rate (dashed line) by age (5 to 18) in the 1881 Population Census of England, for male individuals residing in the county of London. An individual is considered in school if the census records their occupation as "scholar," or in the labor force if any other gainful occupation (i.e. valid HISCO code) is recorded in the census. Results are similar in the 1891 census.

Figure A8: Density of health deficiency index in population vs. estimation sample

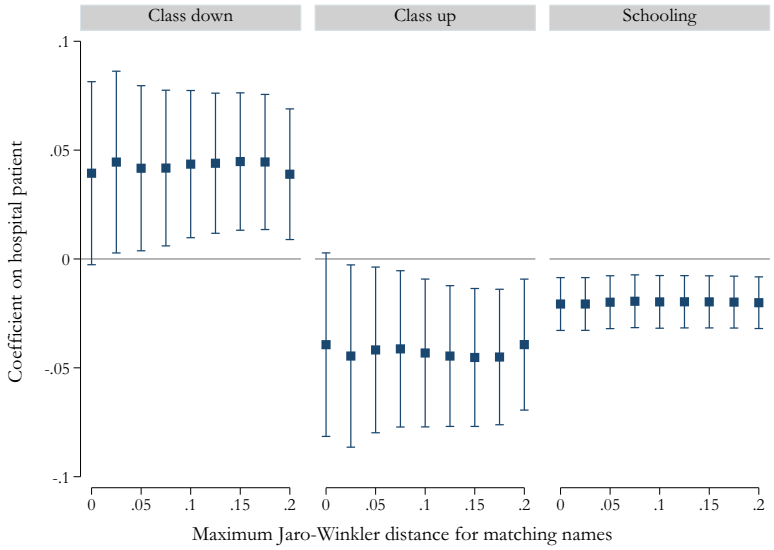


Notes: This figure presents a histogram of the health deficiency index for the population of male patients admitted to the hospitals in the full sample (white) and for the patients in the final estimation sample (solid).

Figure A9: Robustness to changing Jaro-Winkler distance threshold



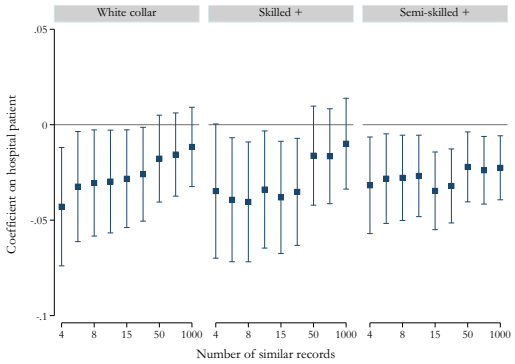
(a) Individual occupational outcomes



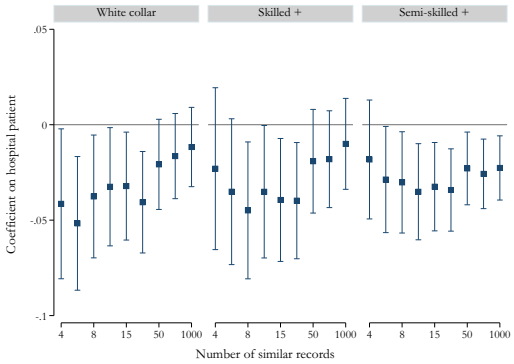
(b) Intergenerational/schooling outcomes

Notes: This figure presents estimated coefficients on the patient indicator variable and 95-percent confidence intervals from separate regressions in which we decrease the Jaro-Winkler distance threshold for inclusion in the sample, moving from the right to the left side of the figures. The left side ($x = 0$) corresponds to the restriction that names must match exactly across censuses and hospital-to-census linkages, while the right side ($x = 0.2$) corresponds to the main sample where Jaro-Winkler distances of up to 0.2 are tolerated. Panel A presents the individual occupational outcomes and panel B presents the intergenerational outcomes and school attendance results. See Table 2 for a description of the empirical specifications.

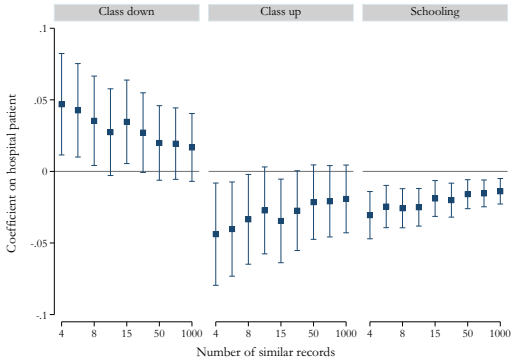
Figure A10: Robustness to changing similar names threshold



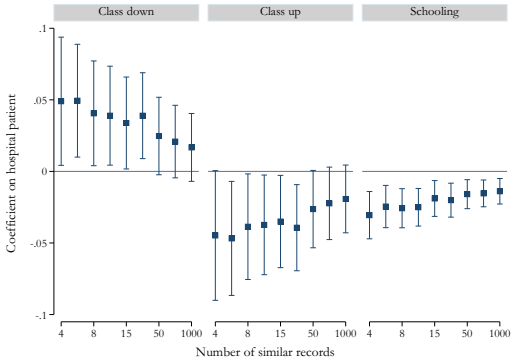
(a) Individual (JW = 0.05)



(b) Individual (JW = 0.10)



(c) Intergenerational/schooling (JW = 0.05)



(d) Intergenerational/schooling (JW = 0.10)

Notes: This figure presents estimated coefficients on the patient indicator variable and 95-percent confidence intervals from separate regressions in which we vary the number of similar names within one year of birth allowed for a uniquely matched record to be included in the sample. Moving from the left to the right side of the figures, we increase the number of similar records allowed from 4 to 1000. In panels A and C, similar names are defined as differing in Jaro-Winkler scores by less than 0.05 for the first and last names compared to the name of the matched individual, while in panels B and D the threshold for the differences in Jaro-Winkler scores threshold is changed to 0.10. The main specification corresponds to the estimates with 20 similar names and a Jaro-Winkler threshold of 0.10. Panels A and B present the individual occupational outcomes and panels C and D present the intergenerational outcomes and school attendance results. See Table 2 for a description of the empirical specifications.