

Karbownik, Krzysztof; Wray, Anthony

**Working Paper**

## Educational, Labor-Market and Intergenerational Consequences of Poor Childhood Health

IZA Discussion Papers, No. 14127

**Provided in Cooperation with:**

IZA – Institute of Labor Economics

*Suggested Citation:* Karbownik, Krzysztof; Wray, Anthony (2021) : Educational, Labor-Market and Intergenerational Consequences of Poor Childhood Health, IZA Discussion Papers, No. 14127, Institute of Labor Economics (IZA), Bonn

This Version is available at:

<https://hdl.handle.net/10419/232879>

**Standard-Nutzungsbedingungen:**

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

**Terms of use:**

*Documents in EconStor may be saved and copied for your personal and scholarly purposes.*

*You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.*

*If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.*

DISCUSSION PAPER SERIES

IZA DP No. 14127

**Educational, Labor-Market and  
Intergenerational Consequences of Poor  
Childhood Health**

Krzysztof Karbownik  
Anthony Wray

FEBRUARY 2021

## DISCUSSION PAPER SERIES

IZA DP No. 14127

# Educational, Labor-Market and Intergenerational Consequences of Poor Childhood Health

**Krzysztof Karbownik**

*Emory University and IZA*

**Anthony Wray**

*University of Southern Denmark*

FEBRUARY 2021

Any opinions expressed in this paper are those of the author(s) and not those of IZA. Research published in this series may include views on policy, but IZA takes no institutional policy positions. The IZA research network is committed to the IZA Guiding Principles of Research Integrity.

The IZA Institute of Labor Economics is an independent economic research institute that conducts research in labor economics and offers evidence-based policy advice on labor market issues. Supported by the Deutsche Post Foundation, IZA runs the world's largest network of economists, whose research aims to provide answers to the global labor market challenges of our time. Our key objective is to build bridges between academic research, policymakers and society.

IZA Discussion Papers often represent preliminary work and are circulated to encourage discussion. Citation of such a paper should account for its provisional character. A revised version may be available directly from the author.

ISSN: 2365-9793

**IZA – Institute of Labor Economics**

Schaumburg-Lippe-Straße 5–9  
53113 Bonn, Germany

Phone: +49-228-3894-0  
Email: [publications@iza.org](mailto:publications@iza.org)

[www.iza.org](http://www.iza.org)

## ABSTRACT

---

# Educational, Labor-Market and Intergenerational Consequences of Poor Childhood Health

Does childhood health capital affect long-run labor market success? We address this question using inpatient hospital admission records linked to population census records. Sibling fixed effects estimates indicate that in comparison to their brothers, boys with health deficiencies were more likely to experience downward occupational mobility relative to their father's occupational rank. This decline in occupational success across generations can be decomposed into a lower likelihood of attaining white collar status and a higher likelihood of working in unskilled jobs, which translated into lower occupational wages on average. Evidence indicates that a lower school attendance rate and higher rates of disability in both childhood and adulthood are plausible mechanisms for our findings.

**JEL Classification:** I14, J62, N33

**Keywords:** childhood health capital, intergenerational inequality, occupational success

**Corresponding author:**

Krzysztof Karbownik  
Department of Economics  
Emory University  
Rich Memorial Building  
1602 Fishburne Drive  
30307 Atlanta, Georgia  
USA

E-mail: [krzysztof.karbownik@emory.edu](mailto:krzysztof.karbownik@emory.edu)

# 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 vicious 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 in estimating the socioeconomic consequences of health during the first few years after birth (Almond and Currie 2011a). However, much less is known about the long-run effects of health throughout childhood (Almond et al. 2018). This distinction could be important to the extent that shocks experienced early in life, such as prenatally, could be more consequential than those that occur later on, for example, at age 10 (Heckman 2006). Notwithstanding differences in the human capital production function, it is also the case that the post-infancy period of childhood is characterized by a relatively lower disease burden (Wang et al. 2020), and thus it is not implausible to think that a child’s health stock in this period is particularly robust. On the other hand, if health matters post-infancy, it may be an important piece of the puzzle missing from explanations for socioeconomic inequality. 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 likewise 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, and lead to long-run adversities (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 affected school enrollment, disability during childhood and adulthood, and long-run occupational success 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 to not only medium-run health and educational consequences, but also long-run occupational outcomes. And third, we combine these inputs with a sibling fixed effects identification strategy.

We construct a multidimensional measure of childhood health deficiencies by collecting historical inpatient records for the universe of children aged 0 to 11 who were admitted to three large London-area hospitals between 1870 and 1902.<sup>1</sup> We refer to our measure as multidimensional to

---

<sup>1</sup>Other papers using hospital inpatient data to study the consequences of poor health include Arthi and Schneider (2020) and Doyle et al. (2020) in historical settings, and Currie et al. (2010) and Schwandt (2018) in modern contexts.

highlight the wide-ranging causes of admissions to the hospital and provide a contrast to the single dimension of an intervention-based measure. We then combine the individual-level hospital records with linked census samples to document the medium-run and long-run effects. The census records contain demographic, educational, and health characteristics from childhood as well as health and socioeconomic outcomes during adulthood. We construct linked samples using complete-count records for the censuses of England between 1881 and 1911, which are newly available from 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. \(2020\)](#). 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](#)). While we are unable to link girls from the hospital records to long-run outcomes due to name changes at marriage, we can observe educational and health status in the medium-run for both genders. In the empirical analysis of long-run outcomes, 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. We find that compared to their brothers, boys admitted to the hospital before the age of 12 were 2.8 percentage points more likely to experience downward occupational mobility, which accounts for 11.3 percent of overall downward mobility in England at the end of the nineteenth century. We also find a decrease of equal magnitude in the likelihood of upward mobility among the affected males. These two results can be decomposed into a decline of 4.3 percentage points in white collar employment, which offsets 17.3 percent of the advantage of having a father in a white collar occupation, as well as an increase in unskilled employment of 3.3 percentage points. Therefore, the effects of childhood health deficiency appear to be concentrated in the upper and lower quartiles of the occupational outcomes distribution, and on average, lead to a 3.1 percent reduction in occupational wages. On the other hand, we do not find any statistically significant effects on long-run social outcomes such as the probability of being married or having children. These results are robust to a variety of specification checks including concerns related to selective mortality, sample selection choices, and changes in the linking algorithm.

We then consider potential mechanisms for the long-run effects and show that hospitalized boys were 2.9 percent less likely to attend school compared to their non-hospitalized brothers in the 10-year period after the hospital admission. Furthermore, the likelihood that boys reported

a disability in the census following admission to the hospital more than doubled in both childhood and adulthood. 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.<sup>2</sup> However, we can ask how poor childhood health affected school enrollment and childhood disability for hospitalized girls compared to their sisters, and find that the effects for both outcomes are similar to those for boys.

We also provide evidence that our results are not explained by pre-existing differences in health status prior to hospitalization or by changes in parental socioeconomic status around the time of hospitalization that differentially affected siblings. Additionally, if the siblings of the children observed in the hospital registers were admitted to hospitals with no surviving records or if they were otherwise unhealthy, then our estimates should be interpreted as lower-bound effects, but we provide suggestive evidence that our results are largely unaffected even if we assume that a non-trivial share of the siblings were hospitalized. However, a limitation of the sibling fixed effects approach is that it cannot separately account for any time-varying factors that are related to the treatment such as compensatory or reinforcing behavioral responses by parents. 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.

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; Warren et al. 2012; Beach et al. 2018; Hoehn-Velasco 2020).<sup>3</sup> Our study is the first to show that deficient childhood health, when proxied by hospitalization, has consequences that extend to the prime earning years and exceed the effects on increased welfare participation in young adulthood found by Currie et al. (2010). Similar to our findings, Hoehn-Velasco (2020) estimates that exposure to a particular preventative health program has a positive impact on adult earnings. On the other hand, our results suggest that the consequences of poor childhood health, broadly measured, extend to a decline in relative occupational status between generations.

Our study is also one of the relatively few papers that measures childhood health using the disease environment beyond the early-life period and assesses its long-run effects. In particular, Bütikofer and Salvanes (2020) explore the impact of a tuberculosis testing and vaccination

---

<sup>2</sup>This significant omission from our knowledge base is now being remedied by the ongoing Longitudinal, Inter-generational Family Electronic Micro-Database (LIFE-M) Project and other related projects making use of birth and marriage registers (Bailey et al. 2020).

<sup>3</sup>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 2015b; Mill and Stein 2016).

campaign, and [Gensowski et al. \(2019\)](#) use data on childhood hospitalizations for polio to study the effects of physical disability from the quasi-random incidence of paralytic polio.<sup>4</sup> In contrast, our hospital admissions data and historical urban setting 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 growing cities ([Fogel 2004](#); [Floud et al. 2011](#); [Currie and Vogl 2013](#)).

Our hospitalization-based indicator of childhood health has several advantages over measures studied in the prior literature, such as modern-day hospital admissions ([Currie et al. 2010](#)), health insurance policy interventions ([Goodman-Bacon 2017](#)), mortality during the early-life period ([Bleakley 2007](#)), and adult height ([Case and Paxson 2008](#)). First, unlike studies investigating policy changes targeted at particular conditions (e.g., antibiotics) or populations (e.g., uninsured), we can be sure, by virtue of data on hospitalization, that treated children in our sample suffered from having a lower health stock. Furthermore, compared with contemporaneous settings, our hospitalization-based treatment yields estimates that are, to a larger degree, unconfounded by the positive effects of inpatient care because medical treatments and mitigation strategies were much less effective historically. Moreover, in the context of our study, hospital admissions occurred 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 our variation in health status 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 during childhood alone, in contrast to adult height, another commonly used proxy for childhood health, which reflects the net influence of health throughout childhood and adolescence.<sup>5</sup> Lastly, it avoids the issues of recall bias and under-reporting with self-reported measures of health ([Smith 2009](#); [Warren et al. 2012](#)).

Our paper also connects to a literature that uses the disease environment around the time of birth as a proxy for early-life health and evaluates its long-run consequences. Studies have typically exploited 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

---

<sup>4</sup>Others have evaluated policy changes during childhood, e.g. expansions in childhood health insurance coverage ([Goodman-Bacon 2017](#); [Brown et al. 2020](#)) and improvements in living conditions ([Gould et al. 2011](#)). The aforementioned [Hoehn-Velasco \(2020\)](#) focuses on the roll-out of county-level health departments in the U.S.

<sup>5</sup>[Case and Paxson \(2008, 2010\)](#) and [Parman \(2015b\)](#) use height as a proxy for childhood health and find that it is positively associated with cognitive test scores and 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.



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 2020) or public health campaigns that sought to eradicate tropical diseases (Bleakley 2007, 2010; Venkataramani 2012; Baird et al. 2016).<sup>6</sup> As these studies define exposure to policy interventions or access to health care at the cohort-by-location level, rather than at the individual level, the first stage effect of treatment on health status is not observed and reported estimates are intention to treat effects. In contrast, our study examines a wide range of childhood health conditions and produces estimates that we can interpret as the average treatment effect of hospitalization, which represents a multidimensional proxy for lower childhood health capital. By observing who is treated, we know which children are likely to have had lower health status. Our results on intergenerational occupational 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, they indicate that poor childhood health can account for a non-trivial share of overall mobility by reducing upward 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, whooping 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, environmental pollution, 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 childcare, 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

---

<sup>6</sup>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. 2020). 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 operating costs out of private endowments, the admission of infants was discouraged on the basis of the beliefs that babies would suffer the ill-effects of separation from their mothers, 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 in part 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 also 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. 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 because 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 establish suitability for admission as an inpatient.<sup>7</sup> Parents of children who were not admitted likewise received some medical advice (Hawkins and Tanner 2013). Hospital admission policies typically recommended the exclusion of 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

---

<sup>7</sup>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).

hospital. However, as we observe in our data, many such cases were admitted in practice.

Despite improved diagnostics that followed gains in knowledge about bacteriology and the increased prevalence of autopsies during our study period, pediatrics was largely limited to convalescent care. 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, cleft lip, or club foot (talipes). For example, 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, the operation to repair cleft lip was considered relatively simple and safe (Lomax 1996). Despite these improvements in surgical treatments, it is not clear whether any long-term socioeconomic disadvantage from scarring or deformation due to the congenital conditions was entirely eliminated.

### **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, disability, and school attendance during childhood, as well as family formation, disability, and occupational 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.<sup>8</sup> 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 the fact that admission records have survived for only four out of the ten largest general hospitals and two of the five largest children’s hospitals in London.

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 A1 shows that Barts and Guy’s Hospitals accounted for 25 percent of 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 or even higher market shares in terms of outpatients and hospital beds.

The hospital records contain detailed information about the patient and the admission. Figure A2 presents a sample page from 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 A2 lists the most common causes of admissions among all patients, all male patients, and those in the sample of male patients used in the main analysis.<sup>9</sup> These included abscesses, diphtheria, pneumonia, chorea, and bronchitis. 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).<sup>10</sup> Additionally, the average mortality rate was relatively constant over the entire sample period (Figure A4). To measure the severity of the insult to health capital, we construct a gender-specific childhood health deficiency index that is based on the unexplained portion of the

---

<sup>8</sup>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.

<sup>9</sup>Throughout the paper we present results separately by gender.

<sup>10</sup>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 at that hospital between 1890 and 1899 was lower at 5.9 percent (Cullis and Young 2013).

likelihood of inpatient death 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 multidimensional childhood health capital that varies across causes of admission.

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.<sup>11</sup> 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 hospitals did not admit patients at older ages.<sup>12</sup> Furthermore, children age 12 and older were much less likely than younger individuals to be living with their parents (Figure A6), which is a necessary condition to implement 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 multidimensional childhood health capital on school attendance, disability during childhood and adulthood, social outcomes such as marriage and fertility, and long-run labor market outcomes 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.<sup>13</sup> 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 and disability status of patients and siblings, and the places of birth (parish and county for domestic births 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

---

<sup>11</sup>The number of observations is balanced across age groups and admission periods with the exception of 1870 to 1875, years for which we restrict to individuals born from 1870 onward (Figure A5).

<sup>12</sup>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.

<sup>13</sup>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 obtained 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).

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.<sup>14</sup> We also create a separate hospital-to-census linked sample of both males and females to explore the role of schooling and childhood disability as potential mechanisms, which we introduce in Sections 3.6 and 3.7.

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 characteristics 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 Ferrie (1996), Abramitzky et al. (2012), Mill and Stein (2016), and Feigenbaum (2016).<sup>15</sup> Appendix A describes the hospital admission-to-census and census-to-census linkage procedures, including how we consolidate multiple hospital-to-census matches and arrive at a final set of unique patient identifiers and individuals in the sample.

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.

---

<sup>14</sup>In an earlier version of this paper (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.

<sup>15</sup>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).



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 causes of admission or occupational outcomes, our final empirical sample includes 2,146 hospital admissions or 7.5 percent of the baseline sample.

### 3.3 Occupational titles, HISCO, and HISCLASS

Our goal in this paper is to understand how childhood health affects long-run labor market outcomes, which we proxy using multiple measures of occupational attainment. Here, we describe the source material for the dependent variables of interest and explain how we translate occupational titles reported in census records into measures of occupational success. In particular, we define and interpret the relative and absolute mobility outcomes that are used in the econometric analysis, and explain why these two sets of outcomes represent different but complementary summary measures of occupational mobility (Section 3.3). We then present descriptive evidence of differences in patient and sibling outcomes in the form of occupational transition matrices in Section 3.4. In Section 3.5 we discuss the construction of occupational wages as an alternative measure of occupational success. Finally, we introduce the data sources for the schooling and disability variables used in the analysis of potential mechanisms (Sections 3.6 and 3.7).

The raw data measuring an individual’s occupational and social rank are the occupational titles reported in the complete count files of the population censuses of England. To each of the text strings, the I-CeM project (UK Data Archive 2014) has assigned the corresponding numeric code from the Historical International Standard Classification of Occupations (HISCO), which is a classification scheme developed for coding nineteenth century occupational titles (Leeuwen et al. 2002). The Historical International Social Class Scheme (HISCLASS) maps each of the 16,000 HISCO occupation codes to one of 12 social classes ranked from highest to lowest 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 (van Leeuwen and Maas 2011). We consolidate the HISCLASS ranking into four groups, 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).<sup>16</sup> Table A3 shows the mapping from the 12-

---

<sup>16</sup>Abramitzky et al. (2011) use a consolidated HISCLASS ranking with seven social classes. In comparison to our four-class ranking, 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

category HISCLASS scheme to the 4-group ranking used in this paper and Table A4 lists the most common occupations in each of the four groups. To help fix ideas, the reader may choose to think about the white-collar, skilled, semi-skilled, and unskilled groups as roughly corresponding to the 1st through 4th quartiles of the occupational class distribution. The most common occupational titles in the four respective groups were clerk, carpenter, carman, and general laborer.

We define two sets of outcome variables for patients and siblings based on data from the 1901 and 1911 censuses, while the occupational rank of their fathers is obtained from the same information in censuses 10 to 30 years prior (i.e. in one of the 1881, 1891, or 1901 censuses). The first set of outcomes includes indicators for “upward” and “downward” mobility which are equal to one if an individual’s occupational rank was higher or lower, respectively, than his father’s rank. These outcomes measure cases where there is an improvement or deterioration in occupational rank across all combinations of father and sons’ ranks. They also have a normative interpretation in that they assume that the stratification of occupations into 4 groups represents an ordered ranking. Arguably, knowing the direction of a change in occupational rank is more informative than defining mobility as a difference in rank in comparison to one’s father regardless of its direction, as has been done in prior work (Long and Ferrie 2013; Pérez 2019). Nonetheless, a limitation of the relative mobility outcomes is that they do not distinguish between differences in the degree of upward or downward mobility. Furthermore, irrespective of the father’s occupational status, attaining white collar status is clearly qualitatively and economically different from entering a semi-skilled occupation. In other words, we also care about an individual’s absolute outcome, and thus we define a second group of outcomes that consist of indicators for a patient or sibling’s own occupational rank in adulthood. Specifically, we construct indicators for white collar occupational rank, white collar or skilled status, and unskilled occupational status. In contrast to the relative mobility outcomes, these outcomes separately measure the absolute likelihood of success at different parts of the occupational distribution. Thus, the relative and absolute outcomes capture different aspects of occupational status, and therefore we view the two sets of measures as complementary.

### 3.4 Occupational transition matrix

Prior to formalizing our econometric analysis, we provide a descriptive comparison of patient-sibling differences in outcomes using occupational transition matrices to summarize the distribution of the father’s occupational rank in comparison to all sons, and to sons separated by their group ranked between skilled and semi-skilled workers. Third, low-skilled and unskilled farm workers are ranked below other unskilled workers. Given that our sample is predominantly urban, we have very few farm operators and unskilled agricultural workers from classes four and seven, respectively. For reasons explained in section 5.3, we also observe few children of professionals among the hospital patients, and thus we choose to consolidate the seven class scheme into four groups.



treatment status (Long and Ferrie 2013; Feigenbaum 2018; Pérez 2019). Table A5 presents occupational transition matrices for households in our main estimation sample. Each column represents the occupational ranks of the fathers (white collar, skilled, semi-skilled, and unskilled) and each row represents the ranks of patients (panel A), control siblings (panel B), patients and siblings combined (panel C), and a synthetic population of linked fathers and sons in England (panel D). The synthetic population is constructed by sampling from five linked samples (1881-1901, 1891-1901, 1881-1911, 1891-1911, and 1901-1911) with sampling probabilities corresponding to the shares of each linked sample in the main estimation sample. Within each panel, a cell reports the percentage of sons in a given rank conditional on father’s occupational rank, and thus the percentages sum to 100 across the four rows in each column. The diagonal elements represent cases in which the father and son are observed to have the same occupational rank, while elements above the diagonal are cases of “upward” mobility in which sons have a higher rank than their father, and elements below the diagonal represent “downward” mobility. For example, among patients in panel A, 45.8 percent of children whose father worked in a white collar occupation also worked in a white collar occupation themselves, but 11.0 percent ended up in unskilled occupations despite their father’s high occupational rank. For control siblings in panel B, the corresponding figures are 49.7 percent and 9.0 percent, respectively, suggesting that poor childhood health might be linked, at least descriptively, to downward occupational mobility. Panels C and D provide descriptive measures of mobility in our main estimation sample and a benchmark population, respectively, which we use to assess the magnitude of our main results in Section 5.

The objective of our long-run analysis is to estimate differences in occupational outcomes between patients and siblings. However, a simple comparison of outcomes based on the transition matrices in panels A and B in Table A5 is potentially problematic, given that patients and siblings will naturally differ on observable as well as unobservable dimensions such as age or birth order. Ideally, we would want to attribute differences between the occupational transition matrices for fathers vs. patients (panel A) and fathers vs. control siblings (panel B) to the causal effect of differences in childhood health, but these comparisons can be confounded. In addition, it’s not clear in this context whether a higher degree of mobility is necessarily a desirable outcome as it does not distinguish between upward and downward mobility, with the former arguably being socially desirable while the latter being detrimental from an individual’s perspective. Therefore, in Section 4, we introduce a regression-based approach with a sibling fixed effects research design to account for those issues.

### 3.5 Occupational wages

While individual-level wage data are unavailable for England during the period of study, Williamson (1980, 1982) produced estimates of occupational wages for twenty-one occupational

categories by census year.<sup>17</sup> We identify individuals who had occupational titles that belonged to one of the occupational categories based on the occupation’s HISCO code and the I-CeM “Occode” variable in the complete-count census records for England and Wales from 1881 to 1911. Similar to Long (2006) and Long (2013), we assign the average occupational wage of an individual’s HISCLASS rank if their occupation does not fall into one of the occupational wage categories.<sup>18</sup> The occupational wages are consistent with our classification of occupational rank into four groups, as the average occupational wage among white collar occupations is 87 percent higher than skilled occupations, while occupational wages in semi-skilled and unskilled occupations are 9 percent and 40 percent lower, respectively.

### 3.6 School attendance

We consider the role of human capital accumulation as one potential mechanism that accounts for the long-run effects on occupational status by constructing a measure of school attendance for children ages 5 to 10. These individuals were subject to compulsory schooling under the Elementary Education Act of 1880.<sup>19</sup> In the absence of specific questions on education in the historical censuses of England, we use information contained in the occupational fields 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 four who attended school on a daily basis. Furthermore, since enumerators typically did not record an occupation for children younger than 12 years old, being regarded as a “scholar” arguably provides a reliable indicator of school attendance.<sup>20</sup>

---

<sup>17</sup>The Williamson (1980; 1982) wage estimates have been criticized (Feinstein 1988) on the basis of the sources used to assign wages for professions such as solicitors and barristers. These concerns have limited significance for our study given that wealthy professionals were rarely admitted to hospitals in nineteenth-century London and our estimation sample only includes three households in which the father’s occupational title was a solicitor or barrister. We address these concerns by replacing the Williamson (1980, 1982) wage estimates for solicitors and barristers, which are outliers in the data, with the average occupational wage in the highest HISCLASS category. This adjustment has no substantive bearing on the estimation results.

<sup>18</sup>Long (2006) and Long (2013) use an alternative classification of occupational titles developed by Armstrong (1972) based on the Registrar General’s 1921 and 1951 classification schemes. We choose to use the HISCLASS ranking to be consistent with the other analyses throughout the paper.

<sup>19</sup>Milner (2020) examines the effects of the UK’s 1870 Education Act which first introduced a public education system in England and Wales, though one without mandatory attendance, and he finds improvements in later-life occupational outcomes of affected children. The 1880 extension effectively made the schooling mandatory until age 10. 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.

<sup>20</sup>We code an individual as a scholar if the occupation string from the childhood census contains the words scholar, student, or school. We allow for typographical errors and exclude cases that contain the following words: master, mistress, monitor, monitress, teacher, and under training. We also identify scholars from occupation codes (Occode) and relationship to household head (Rela) variables which include codes for different types of students.

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 were ages 5 to 10 years old at the time of the census.<sup>21</sup> 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.

### 3.7 Disability

Another mechanism for our long-run occupational findings is the persistence of poor health. Empirically, we examine this idea using a measure of extreme health impairment – disability status as reported in the census. Individuals were asked about their disability status in all censuses between 1881 and 1911, which allows us to study effects on disability in both childhood and adulthood. While the raw census data contain a string variable for the disability recorded by the enumerator, our measure is based on two enriched variables coded by the I-CeM project. The first assigns disabilities to one of five categories (DisCode1): visual impairment, hearing impairment, idiocy and imbecility, lunacy, or other disability. The second (DisCode2) contains more detailed information on physical disabilities. We code disability as an indicator variable taking the value of one if an individual reports *any* disability in one of the categories for either of the two aforementioned variables.

For the childhood disability analysis, our sample consists of male and female patients linked from the hospital records to one of the 1881 through 1911 censuses up to 10 years post admission, as well as their siblings located in the same household in the census. In the same manner as in the main occupational analysis, we restrict the sample to patients born between 1870 and 1890 and admitted at ages 0 to 11. We further require that both patients and siblings be 21 years or younger at the time of census enumeration. When we turn to disability in adulthood we can only include men in the sample due to name changes at marriage by women. The construction of the long-run sample follows the procedure outlined for the long-run occupational outcomes, but we additionally include individuals with missing occupational outcomes. In a separate exercise, we examine whether hospitalization is associated with within-household differences in the likelihood

---

<sup>21</sup>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 census for this exercise.

of disability prior to admission. In this case, the sample construction is identical to the childhood disability sample, with the exception that patients are linked from the hospital records to a census up to 10 years before admission.

## 4 Empirical Specification

As we have foreshadowed in the previous section, our main empirical strategy involves a sibling fixed effects research design. Here, we formalize the assumptions needed to identify the causal effect of multidimensional childhood health deficiencies on long-run labor market outcomes. We also build upon our discussion of the differences in the interpretation of the relative and absolute occupational outcomes from Section 3.3 and illustrate how sibling fixed effects estimates will differ across specifications with either set of outcomes.

### 4.1 Potential outcomes

Let  $D_i \in \{0, 1\}$  denote whether or not an individual is treated, where  $D_i = 1$  corresponds to the case where an individual is hospitalized. Thus, an individual is either a hospitalized sibling or a control sibling. Potential outcomes in the untreated and treated states are denoted by  $Y_i(0)$  and  $Y_i(1)$ , respectively. For each individual  $i$  in the data, we observe a household  $H_i$ , treatment status  $D_i$ , and outcome  $Y_i = Y_i(D_i)$ . The causal effect of the treatment for an individual  $i$  is  $\tau_i = Y_i(1) - Y_i(0)$  and the household-level treatment effect is:

$$\tau_{H_i=j, FE} = \mathbb{E}[Y_i(1)|H_i = j] - \mathbb{E}[Y_i(0)|H_i = j]$$

Since  $Y_i(0)$  is not observed for individuals who are treated, we make the following conditional independence assumption:<sup>22</sup>

$$Y_i(0), Y_i(1) \perp\!\!\!\perp D_i | H_i = j$$

This assumption allows the hospitalization treatment to be correlated with household characteristics, but requires it to be as good as randomly assigned within the household conditional on household fixed effects. As noted above in Section 3.1, we interpret hospitalization as a proxy for childhood health deficiency. Thus, the conditional independence assumption implies that within-family variation in multidimensional health capital is uncorrelated with unobserved child-specific characteristics that determine occupational success. Thus, we can re-write the household-level

---

<sup>22</sup>For simplicity and for illustrative purposes we omit control variables from this derivation but include them in the econometric specification described in Section 4.2.

treatment effect as:

$$\begin{aligned}
\tau_{H_i=j,FE} &= \mathbb{E} [Y_i(1)|H_i = j] - \mathbb{E} [Y_i(0)|H_i = j] \\
&= \mathbb{E} [Y_i(1)|D_i = 1, H_i = j] - \mathbb{E} [Y_i(0)|D_i = 0, H_i = j] \\
&= \mathbb{E} [Y_i|D_i = 1, H_i = j] - \mathbb{E} [Y_i|D_i = 0, H_i = j]
\end{aligned}$$

When the CIA holds, the estimated treatment effects  $\hat{\tau}_{H_i=j,FE}$  will be an unbiased estimator of the treatment effect of interest.

## 4.2 Sibling fixed effects estimator

We implement a sibling fixed effects research design to deal with concerns that the hospitalization treatment is correlated with time-invariant household characteristics that are also determinants of the outcomes. We identify the effects of childhood health deficiencies on school attendance, disability, and occupational success for male patients and their brothers by estimating the following regression:

$$Y_{ij} = \alpha_j + \beta D_{ij} + \mathbf{X}_{ij}'\gamma + \varepsilon_{ij} \quad (1)$$

where  $i$  indexes individuals and  $j$  indexes households. Our data structure implies that we observe information for  $i \in \{1, \dots, N\}$  individuals who belong to household  $H_i = j$  for  $j = 1, \dots, H$ . We refer to the set of individuals in a household ( $i \in H_i$ ) as siblings. In general, the number of observations in each household  $j$  can range from 2 to  $N$ , but to facilitate a clear interpretation of our estimates, we ensure that our samples are always balanced so that each household  $j = 1, \dots, H$  includes exactly two individuals: a patient and the sibling closest in age who remains in the data set after implementing our matching procedure. Given that exactly half of our sample is hospitalized, our estimate of the treatment effect can roughly be interpreted as the average treatment effect of hospitalization (Słoczyński 2020). We test the consequences of this restriction in Section 6 and show that our results are unchanged when it is relaxed.

The variable  $Y_{ij}$  represents either one of the binary outcomes for relative or absolute occupational rank that were defined in Section 3.3 and which vary across the specifications, or occupational wages, while  $\alpha_j$  denotes a sibling fixed effect that captures unobservable time-invariant factors common to the household. In the primary specifications, we measure childhood health deficiencies using an indicator for hospitalization  $D_{ij}$  which is equal to one for patients and zero for control siblings. The equivalent assumption to the CIA for potential outcomes is the strict

exogeneity assumption for FE estimators:

$$\mathbb{E} \left[ \varepsilon_{ij} | D_{ij}, \alpha_j, \mathbf{X}'_{ij} \right] = 0$$

which implies that conditional on household fixed effects  $\alpha_j$  and controls  $\mathbf{X}_{ij}$ , the error term in the sibling FE regression  $\varepsilon_{ij}$  is orthogonal to treatment status  $D_{ij}$  for both patients and siblings. In other words, childhood health deficiency is as good as randomly assigned amongst siblings conditional on controls. The coefficient of interest is  $\beta$ , which can be interpreted as the differential effect of lower childhood health capital for a hospitalized child compared to his brother. If the strict exogeneity assumption holds, then  $\beta$  represents the average treatment effect  $\tau_{ATE}$ . The sibling fixed effects estimator for  $\beta$  is only identified from households in which the binary treatment is different for patients and siblings. In our data, all households have variation in treatment status by construction, and thus the concerns raised by Miller et al. (2019) about selection into identification related to households that have variation in the treatment do not apply to our empirical setting.<sup>23</sup> The vector of individual characteristics  $\mathbf{X}_{ij}$  includes age at enumeration-by-census year fixed effects, birth order fixed effects, an indicator for the older sibling in the pair, measures of name commonness to proxy for the SES content of names (a standardized measure of first name frequency and an interaction of first name and surname frequency), and matching controls (indicators for exact matches on first name, surname, and birth parish between censuses, as well as for cleaned birth places).<sup>24</sup> An error term  $\varepsilon_{ij}$  is clustered at the childhood household level and represents sibling-specific unobserved characteristics. In separate specifications, we replace the indicator for hospitalization with the health deficiency index that is described in appendix B.

Next, we provide an illustrative example to highlight how the sibling fixed effects estimator for the effect of childhood health deficiency will differ in specifications with the relative mobility outcomes compared to those with the absolute occupational outcomes. Recall that we introduced these outcomes in Section 3.3 and highlighted the differences in their respective interpretations as the former capture changes in status relative to one’s father across the occupational distribution, whereas the latter measure the likelihood of attaining a rank at a specific part of the occupational distribution. Now, consider two households in which the father is observed in an unskilled occupation. In the first household, the healthy sibling is observed in a skilled occupation, while the

---

<sup>23</sup>Additionally, selection into identification is not a concern in our setting because throughout the paper we use linear probability models. Non-linear models such as fixed effects logit require “double switchers” for identification, which means they also exclude families where there is no variation in the outcome variable. This means that moving across the two types of models will change the sample size and effectively the set of families identifying our average treatment effect.

<sup>24</sup>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.



hospitalized patient enters an unskilled occupation similar to his father. In the second household, the hospitalized patient ends up in a semi-skilled occupation while the healthy sibling enters a white collar occupation.

If we consider the first household, the relative outcome of upward mobility will differ between patient and sibling, since only the sibling moves up in status relative to his father, whereas in the second household there is no difference in relative mobility between patient and sibling, since both improve in occupational rank compared to their father.<sup>25</sup> The latter case highlights a limitation of the relative mobility outcome in that it does not distinguish between differences in the degree of upward mobility, i.e. rising by two social ranks is better than rising by one rank. Nonetheless, we include the relative mobility measures since the interpretation of these outcomes is comparable to the analysis of transition matrices that has been explored extensively in the occupational mobility literature (Long and Ferrie 2013; Feigenbaum 2018; Pérez 2019). Regardless, we also examine absolute mobility outcomes such as an indicator for white collar occupational status. For the second household in the example, there is a difference in absolute outcomes between patient and sibling, given that ending up in the middle of the occupational distribution is worse than rising to the top. However, in the first household, there is no difference in white collar status between patient and sibling, as neither attains white collar status.<sup>26</sup> As the example illustrates, in specifications with a relative mobility outcome, a different set of households contribute non-zero differences to the sibling fixed effects estimator compared to specifications with the absolute occupation outcomes. Thus, the estimates will differ and capture complementary but distinct measures of occupational success.

## 5 Main Results

We begin our discussion of the results by analyzing the effects of childhood health deficiencies on long-run occupational success using the main estimation sample of male patients and their brothers linked to census records during adulthood. We put the magnitudes of these estimates into context by estimating intergenerational elasticities for occupational status and wages,

---

<sup>25</sup>Importantly, the coefficients on upward and downward mobility are not necessarily of equal magnitude and of opposite sign. Households in which one sibling moves up in status and the other moves down relative to the father will have differences between patients and sibling in both the upward and downward mobility outcomes. Yet, a household in which one sibling moves up and the other stays in the same occupation as their father will have positive upward but zero downward mobility, while a household with one sibling moving down and the other staying in the same occupation as their father will have positive downward but zero upward mobility.

<sup>26</sup>In the final empirical sample, 25 to 28 percent of households have variation in the mobility measures while 20 to 39 percent have variation in the own rank outcomes. An indicator for households with patient-sibling pairs that have variation in outcomes is uncorrelated with potentially confounding explanatory variables such as 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.

making comparisons to previous findings in the literature, 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 and the persistence of health deficiencies by considering school attendance and disability as intermediary outcomes, respectively. Lastly, we examine the heterogeneity in the effects by cause of admission and age at admission, as well as the external validity of our findings.

Our main sibling fixed effects regression results for the long-run occupational outcomes are presented in Table 2. Panel A presents estimates in which the treatment of interest is an indicator variable for hospitalization, while panel B replaces the treatment with the health deficiency index that exploits variation in severity across causes of admission. Columns 1 and 2 display the results for occupational success relative to the father's rank, while columns 3 to 5 decompose these mobility outcomes into the likelihoods of attaining a particular occupational rank: white collar, skilled or white collar, and unskilled. The final column presents results with log occupational wages as an alternative measure of labor market success.

We find that poor childhood health, when proxied by hospitalization, reduces the likelihood that a patient, relative to his brother, surpasses his father's occupational rank. This 3.0 percentage point (pp) reduction corresponds to a 8.5 percent effect relative to the mean. Patients are also 2.8 pp or 10.6 percent more likely than their siblings to experience a decline in occupational success relative to the socioeconomic status of their childhood household. Taken together, these results indicate that the occupational status of patients was worse than their non-hospitalized siblings in the long-run. While the specifications with the upward and downward mobility outcomes have a convenient normative interpretation, they do not reveal the parts of the occupational distribution to which patients and siblings were more or less likely to belong. In addition, the specifications treat the outcomes of patients and siblings as being the same in cases where they attain different status levels that are both higher than their father's, which arguably represents useful variation. Thus, to unpack our directional mobility results, we next consider outcomes that account only for the absolute rank of patients and siblings during adulthood.

Our findings for the own occupational rank outcomes indicate that the effects on upward and downward mobility are driven by patient-sibling differences in the tails of the occupational distribution. In particular, we find that poor childhood health reduces the probability of attaining a white-collar occupation as an adult by 4.3 pp or 15.9 percent relative to the mean. The estimates for the likelihood of entering skilled or white-collar occupations, and for ending up in an unskilled occupation, are 6.7 and 21.5 percent, respectively. On the whole, hospital patients were less likely than their brothers to move up to the highest professional ranks and more likely to fall into the lowest status positions as a consequence of their poor health during childhood. These movements



across the occupational rank distribution lead to occupational wages that are 3.1 percent lower, on average, for children with lower health capital than their brothers. Taken together, the long-run occupational results reveal that patients, on average, experienced worse occupational mobility than their siblings in both relative and absolute terms.

Given that the binary measure of hospitalization treats all causes of admission equally, we also present estimates with the health deficiency index (HDI) as an alternate treatment variable, the construction of which we discuss in Section 3.1 and appendix B. The HDI exploits the variation in severity across causes of admission and is standardized on a 0 to 1 scale with higher values indicating more severe diagnoses. In the population of hospital patients, the mean value of the HDI is 0.30 and the standard deviation is 0.13. To fix ideas, a one-standard deviation increase in the HDI relative to the mean corresponds to being admitted for heart disease (*morbus cordis*) or sequela of diphtheria (diphtheric paralysis) as opposed to the causes of admission around the mean value of the HDI such as rheumatism and diseases of the hip or knee. A change in severity of this magnitude decreases the likelihood of attaining higher occupational status as an adult by 2.8 to 6.7 percent, depending on the specific outcome variable used. These effects, along with an increase in the probability of ending up in an unskilled occupation by 8.7 percent, translate to a drop in occupational wages of 1.2 percent.<sup>27</sup> In sum, the HDI results highlight the degree to which a modest increase in the severity of the hospitalized condition resulted in lower occupational status for patients compared to their siblings.

While the results for the effects of poor childhood health on long-run occupational outcomes presented thus far hint at effect sizes that are large in absolute magnitude, it is not yet clear whether the effects were economically meaningful in the context of late-nineteenth and early-twentieth century England. To put the magnitude of our estimates into context we conduct three exercises. First, we compare them to the association between father and son’s status in a sample of over 2 million children age 0 to 11 in England linked between the 1881 and 1911 census complete-count files. Second, we make a comparison to previous results in the literature, and third, we consider the overall mobility rates in the population.

In panel A of Table A6 we present results from estimating the following regressions that measure associations of father’s and son’s occupational status:

$$Son's\ status_i = \alpha + \beta \cdot Father's\ status_i + \gamma X_i + \varepsilon_i \quad (2)$$

where  $i$  indexes individuals, and *Son’s status* and *Father’s status* are one of three measures of sta-

---

<sup>27</sup>The standard deviation of the health deficiency index (HDI) is 0.13. A change in the HDI of this magnitude reduces log wages by  $0.096 \times 0.13 = 0.0125$  log points. This change represents a  $100 \times [\exp(0.0125) - 1] = 1.2$  percent change in log wages.

tus as adults: indicators for white collar occupations or skilled occupations and above, or log occupational wages.  $X_i$  is a vector of individual characteristics that includes son's and father's age fixed effects, indicators for a first-born child and above median sibship size, as well as the measures of name frequency and match quality introduced in Section 4.2. Columns 1 and 2 indicate that having a high status father increases the likelihood of attaining high status as an adult by 22 to 25 pp. In column 3, we regress the log occupational wage of the patients and siblings as adults on their father's log occupational wage and estimate an intergenerational occupational wage elasticity of 0.246. This estimate is comparable to the intergenerational elasticity of income estimates based on father's and son's annual log earnings from the state of Iowa between 1915 and 1940 which range from 0.199 to 0.258 (Feigenbaum 2018).

Next, we scale the estimated coefficients on the hospitalization indicator from columns 3, 4, and 6 in panel A of Table 2 by the estimates for the intergenerational transmission of occupational status.<sup>28</sup> These results are presented in panel B of Table A6. We find that on average, health deficiencies in childhood offset 16.5 to 17.3 percent of the intergenerational status premium.<sup>29</sup> When it comes to occupational log wages, the effect of lower health capital offsets 13.0 percent of the intergenerational occupational wage elasticity. We can also scale the effects of a 1 s.d. (0.13) change in the health deficiency index reported in panel B of Table 2 by the estimates for the intergenerational transmission of occupational status (panel A of Table A6). Such a negative health effect offsets about 5.0 to 7.3 percent of the advantage of having a higher status father. These scaled effects are smaller given that a 1 s.d. change in the HDI represents a smaller change in health status than the difference between hospitalized patients and non-hospitalized siblings. Across the various comparisons of magnitudes, the consequences of poor childhood health appear socially and economically relevant in England during the early-twentieth century.

Turning now to the second exercise for contextualizing the magnitudes of our estimates, we highlight their comparability with other findings in the literature. The magnitudes of the scaled effects of hospitalization shown in Table A6 are strikingly similar to the effects of a tuberculosis control program in Norway during the 1950s which reduced the persistence of educational attainment across generations by 14.3 percent in comparison to the pre-intervention levels (Bütikofer

---

<sup>28</sup>For ease of interpretation, we do not show results from column 5 of Table 2 with an indicator for unskilled occupations since it is a measure of lower occupational status, in contrast to the analysis of father-son associations in higher occupational status discussed here. Nonetheless, the results are similar if we replace the outcome for unskilled occupations with an indicator for occupations ranked semi-skilled or higher.

<sup>29</sup>If we take, for example, the estimates for effects on white collar status in column 1, it can be seen that the 17.3 percent scaled effect of hospitalization is slightly larger than the 15.9 percent effect on white collar status in column 3 (panel A) of Table 2. This difference arises from the fact that the estimates presented here are scaled by a coefficient with a magnitude of 25 pp that represents the difference in the likelihood of attaining white collar status for sons of white collar fathers in comparison to sons with fathers whose occupational rank was below white collar status, whereas the own occupational rank estimates in Table 2 are scaled by the share of white collar sons in the main estimation sample (27.3 percent). Ex ante, it is not the case that these effects should be similar.

and Salvanes 2020). Additionally, in complement with results from the U.S., the 3.1 percent reduction in log occupational wages due to hospitalization during childhood (Table 2) is larger in absolute magnitude than the 1.2 percent increase in log wages from eliminating early-life exposure to typhoid fever (Beach et al. 2016) and smaller than the 4.6 percent increase in log wages from exposure to community health departments at ages 0 to 5 for treated siblings in comparison to their brothers (Hoehn-Velasco 2020). These comparisons suggest that our results are broadly representative of impacts on intergenerational and long-run labor market outcomes due to changes in childhood health studied in the context of specific policy interventions or diseases. Moreover, they increase the external validity of the studies focused on singular health shocks.

As a third benchmark for the upward and downward mobility results in columns 1 and 2 of Table 2, we consider the extent of upward and downward occupational mobility in the synthetic linked population presented in panel D of Table A5 and introduced in section 3.4. The rates of upward and downward mobility in the synthetic linked population are somewhat lower than in our estimation sample at 31.8 and 24.7 percent compared to 35.7 and 26.0 percent, respectively. Thus, the scaled effect of the patient indicator accounts for 9.4 percent of overall upward mobility and 11.3 percent of overall downward mobility in the linked population, which likewise suggests that poor childhood health accounts for a meaningful but reasonable share of overall mobility in England at the turn of the twentieth century.

We have also examined effects of childhood health on additional long-run outcomes that capture social rather than economic aspects of life, including marital status, fertility, or migration. These results are presented in Table A7 for both the hospitalization indicator (panel A) and the health deficiency index (panel B). Irrespective of the exact specification, we do not find any statistically significant results for these non-labor market outcomes. The direction of coefficients is consistent with the long-run disadvantage of individuals with poorer childhood health, as the patients are more likely to live with their parents, less likely to move or be married, and have lower fertility as adults. On the other hand, conditional on having at least one child, they are more likely to send them to school (column 5), although this estimate is likewise far from statistical significance. The effect sizes for these estimates are also much smaller than the main results for long-run occupational outcomes reported in Table 2, but we cannot credibly rule out potentially meaningful effects given the wide bounds on the 95 percent confidence intervals. For this reason we focus solely on our main labor market results in subsequent sections.

## 5.1 Mechanisms

Next, in Table 3, we consider potential mechanisms and mediators linking poor childhood health to long-run occupational success. To begin, we inquire whether poor childhood health

limited opportunities for human capital accumulation, which in turn could have impeded outcomes 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. 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.

The school attendance results are documented in columns 1 to 3 of Table 3. The first two columns display results from single sex samples where we compare male patients to their brothers and female patients to their sisters. The third column then shows results from a larger sample that pools together the boys and girls from the single-sex samples in the first two columns to increase our statistical power. As before, we present estimates 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, as hospitalization is associated with a 2.1 to 3.2 pp (2.9 to 4.3 percent) decrease in the probability of attending school up to 10 years after hospitalization. Similarly, a 1 s.d. (0.13) increase in the health deficiency index leads to a 1.4 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, at least when it comes to the patient indicator treatment variable. Thus, we conclude that the reduction in school attendance likely impeded human capital accumulation in childhood 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 columns 1 to 3 of Table 3, we also examine the persistence in the effects on health in both the medium and long run. Columns 7 to 9 present effects on childhood disability observed up to 10 years post admission for the within-gender sibling comparisons. Since severe disability is an extremely rare outcome that is reported by only 0.5 percent of the sample, we multiply these dependent variables by 100 to ease the interpretation. With that caveat in mind, we find very robust differences in childhood disability rates between hospitalized patients and their non-hospitalized siblings across all samples. Since the mean of the dependent variable is very low, all estimated effect sizes exceed 100 percent. These findings suggest that, for at least some patients, the scarring effect of poor childhood health manifested as general disability later in adolescence.

Given the short-run results for childhood disability, it is natural to ask whether these disabilities persisted into adulthood and could have plausibly restricted the pool of occupational choices. For boys only, we can examine these long-run effects in a sample with an additional linkage to a

census during adulthood. Column 10 shows that the differences in the rates of disability between patients and siblings persist into adulthood, at the same range of ages at which we measure labor market outcomes. Thus, these results suggest that the main effects on occupational success could, at least partially, work through the persistence of poor health. Nonetheless, given the low rates of reported disability, these results should not fully explain our occupational findings. However, it is possible that a non-trivial number of individuals who experienced a negative health shock in childhood may have suffered from long-run health deficiencies that were sufficiently serious to impede their occupational success, but were not reported in the census, perhaps due to their lower severity than the extreme conditions recorded by census enumerators, or due to the stigma of disability in Victorian England (Cohen 2014).<sup>30</sup>

While suggestive of persistence in poor health, the disability results could otherwise reflect pre-existing health differences prior to hospitalization that jointly determine the likelihood of being sent to a hospital and being observed with a disability in the census records. In columns 4 to 6 of Table 3, we address this issue by examining the relationship between our two measures of poor childhood health and the likelihood of being observed with a disability in the census prior to the hospital admission for patients compared to siblings. Irrespective of the exact sample and specification, we do not find any statistically significant estimates, but it is important to recognize differences in magnitudes across genders. For boys who can be linked to long-run labor market outcomes and represent the primary focus of this paper, we find a very small point estimate. On the other hand, the point estimate for females is large in relative terms, although it is less than one-fourth the size of the corresponding post-hospitalization childhood disability estimate reported in column 8 of panel A. The results for both genders are more encouraging when using the health deficiency index as the treatment variable, in which case the point estimates for effects on pre-existing disability (columns 4 to 6) are orders of magnitude smaller than those for effects on childhood disability (columns 7 to 9), and in each case we can reject their statistical equality. Given the inconsistency in the magnitudes across the two treatment measures for the effects on pre-existing disability for girls, and that our primary focus is on boys, the large point estimate in column 5 of panel A should not be a concern for the main results in the paper. Overall, we view these results as suggestive that differential disability prior to hospitalization was not an important driver of our measure of childhood health deficiency and later long-run outcomes. They further suggest that our results are not explained solely by differences in prenatal or infant health, or at least not by those leading to disability. Rather, in support of our identifying assumption, it appears that hospitalization could have resulted from an exogenous health shock that affected the

---

<sup>30</sup>The disability results in Table 3 use samples from the same birth cohorts (1870 to 1890) as the long-run occupational results in Table 2, given that we view disability as a mechanism. In appendix Table A8 we show that the disability results are very similar when estimated on an expanded sample that includes patients born as late as 1902.

admitted child but not their sibling, which then lead to a divergence in disability, schooling, and labor market outcomes later in life. The persistence of this health shock, via disability, and reduced schooling are credible but likely not exclusive mechanisms behind this gap in occupational outcomes.

Another possible mechanism that also represents a potential confounder in our analysis is a decline in the occupational status of the father. A negative employment shock experienced by the father could have triggered differences in within-sibling childhood hospitalization due to lower quality of nutrition available to children who experienced the shock during critical periods of growth. These differences could then lead to adverse occupational outcomes for these children. We deal with this concern in two ways. First, we proxy for sibling-specific socioeconomic status (SES) by counting the number of young children (age 0 to 5) who were in the household at two points in time for each individual in our sample: during infancy (age 0 to 2) and at the time of hospitalization. In the latter case, for control siblings we assign the number of children present at the same age as the hospitalized sibling's first admission. In our main specification, we include flexible controls for these variables but they do not alter our coefficients of interest. Second, to the extent that a shock to household SES was permanent, it would be picked up by sibling fixed effects. Irrespective of the fixed effects, it would still be the case that a younger sibling is exposed to lower SES for a longer period of time than an older sibling, but this correlation should be captured by the birth order fixed effects and the indicator for whether the hospitalized child is the older or younger child within a pair of siblings in the sample, which are included in our regression models.

Lastly, we also provide descriptive evidence of potential changes in father's SES based on a sample of fathers linked between subsequent censuses.<sup>31</sup> Figure A8 shows that 70 percent of fathers have the same occupational rank in both censuses, which suggests that a shock to household SES that could differentially affect patients and their siblings is an unlikely explanation for the main results. Approximately 15 percent of fathers experience a reduction in their occupational status and a similar fraction experience an increase across census years. Based on these numbers, we randomly impute the father's SES to be one rank lower for 15 percent of patients in the sample. We repeat this exercise 1,000 times and report the distribution of point estimates for the effects of poor childhood health on long-run outcomes in Figure A9.<sup>32</sup> Regardless of the exact outcome, our results remain very similar to those presented in Section 5. Thus, we conclude that unobserved changes in paternal SES are an unlikely confounder or mechanism.

In short, we find evidence supporting both human capital accumulation and the persistence of

---

<sup>31</sup>We describe the procedure for generating the linked sample of fathers in appendix A.5.

<sup>32</sup>Note that this change directly affects the values of our downward and upward mobility outcome variables. For the remaining outcomes we include flexible controls for father's status in the sibling fixed effects specifications.



poor health as plausible mechanisms for the long-run occupational mobility results. Conversely, we show that our results are not driven by pre-existing health difference prior to hospitalization or by changes in parental SES. Nonetheless, we cannot rule out reinforcement of childhood health shocks by parents as another channel and thus our reduced-form estimates may partly capture the role of parental behavioral responses triggered by their children’s lower health capital.

## 5.2 Heterogeneity

Having presented evidence on the mechanisms for the effects on long-run occupational mobility, we now elaborate further on our interpretation of hospitalization as an indicator for childhood health deficiencies by considering the heterogeneity in the effects of hospitalization by cause of admission. An advantage of using hospital records rather than intervention-specific health shocks to study childhood health is that the causes of admission are wide-ranging, from cases such as birthmarks, crossed eyes, or enlarged tonsils that are not expected to yield detrimental effects, to conditions such as diphtheria that may cause severe consequences in the long-run. However, our main results may not be generalizable if they are driven solely by sub-groups of patients who suffered from particularly severe or chronic health insults that resulted in a persistent scarring and lead to permanent disability. Thus, we modify equation 1 by including separate indicators  $D_{ij}(c)$  for each cause of admission instead of a single hospitalization indicator  $D_{ij}$ :

$$Y_{ij} = \alpha_j + \sum_{c=1}^N \beta_c D_{ij}(c) + \mathbf{X}_{ij}'\gamma + \varepsilon_{ij} \quad (3)$$

We specify the indicators for causes of admission according to three different categorization schemes. First, we assign causes of admission to broad categories: acute, chronic, contagious, injuries, congenital conditions, and other admissions. Second, we estimate separate coefficients for the common causes of admission listed in Table A2. Third, we categorize causes of admission using the body system classification developed by [Kingston University \(2010\)](#). In each case, the set of admission dummies includes an indicator for multiple categories of admission and an indicator that pools together all remaining causes of admissions. The coefficients of interest  $\beta_c$  are interpreted as the effect of suffering from cause of admission  $c$  in comparison to a non-hospitalized sibling. The results are reported in Figures A10 to A12.

Across the three categorizations, we find a substantial degree of heterogeneity by cause of admission. While individual coefficients are for the most part statistically insignificant, F-tests suggest that groups of coefficients for causes of admission with the largest effect sizes are jointly significant. For example, in the specification with white collar occupational status as the outcome, the coefficients on admissions for eczema, chorea, tuberculosis, cleft lip, talipes, and fractures are

jointly significant at conventional levels. It is clearly the case that the results are not driven by a specific type of admission. Figure A10 presents the coarsest classification and shows that the largest effects are found for acute admissions, injuries, and congenital conditions. A finer classification in Figure A11 indicates that causes of admission with the largest effects, albeit separately statistically insignificant, include infectious diseases such as diphtheria, scarlet fever, and typhoid fever, as well as congenital conditions such as cleft palate and cleft lip. While admissions for specific injuries are less common, the effect of fractures is similar in magnitude to the overall effects. Similarly, Figure A12 suggests that a wide range of body systems were affected. We view these results as supporting our interpretation of hospitalization as a proxy for multidimensional childhood health deficiencies. Furthermore, they suggest that the childhood health conditions that lead to long-term consequences extend beyond the infectious diseases such as influenza (Almond 2006; Beach et al. 2018), polio (Gensowski et al. 2019), tuberculosis (Bütikofer and Salvanes 2020), typhoid fever (Beach et al. 2016), and yellow fever (Saavedra 2017), or parasitic diseases such as hookworm (Bleakley 2007) or malaria (Barreca 2010) that have been studied in the existing literature, primarily in the context of exposure during the prenatal period or infancy.

Next, we present sample splits that provide some evidence on two additional sources of heterogeneity: the distribution in the severity of the health insult and the age at admission. Table A9 displays these results, which should be treated as suggestive only since we cannot reject the null of equal coefficients at conventional levels due to smaller sample sizes, even in cases where these estimates appear qualitatively different.

Given that the health deficiency index (HDI) introduced in Section 3.1 provides a measure of severity, in panel A of Table A9 we estimate separate effects for patients admitted for conditions with a HDI value above and below the median index value in the hospital population. We find very similar coefficients for the mobility estimates and the white collar occupational indicator, while the effects appear larger for above-median HDI conditions when we consider lower levels of occupational attainment (columns 4 and 5). These results suggest that despite the potential for positive selection into survival for patients admitted with high mortality conditions, the scarring effect of these conditions is associated with worse occupational outcomes than the effects of less severe illnesses. On the other hand, our occupational wage results suggest that below-median HDI conditions had more detrimental consequences. As can be seen in Figures A10 to A12, some of the largest effects on occupational log wages by cause of admission are congenital conditions as well as diseases of the skin or the urinary system, all of which had low in-hospital mortality. This result highlights a limitation of the HDI as a proxy for severity, since low mortality conditions could also have long-term consequences.

Previous literature suggests that shocks experienced earlier in childhood could be more con-



sequential than those that occur later on (Heckman 2006). Thus, in panel B, we present results in which we estimate separate effects for children who were first admitted to the hospital before age 5 and those who were first admitted at ages 5 to 11.<sup>33</sup> Our point estimates are between 1.4 and 2.9 times larger for children admitted early rather than late in childhood, with the exception of downward mobility for which the estimates are larger for shocks occurring later in childhood. In 9 out of 12 estimates, however, the estimates have the expected sign and are statistically significant. 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.

### 5.3 External validity

Thus far we have described the main results, with effects on intergenerational occupational mobility, occupational rank, and occupational wages; the mechanisms, which include schooling and disability channels; and heterogeneity in the effects of hospitalization by cause of admission. Now we turn to the external validity of the estimates, seeing as our analysis revolves around the selection of individuals into hospitalization.

As we cannot speak to hospitals without surviving records, we 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 A10 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.<sup>34</sup> 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 than 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,

---

<sup>33</sup>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.

<sup>34</sup>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.

or the share of immigrants.

Next, Table [A11](#) 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. When 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. 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 in the privacy of their homes and relied on general practitioners who operated private clinics ([Carpenter 2010](#)). Importantly, 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, and indeed it further motivates our empirical design.

We also examine the influence of differential linkage rates on our main estimates by reweighting the observations in our sample by the observable characteristics of all patients in the hospital records. This procedure is described in appendix [C](#) and draws on [Abramitzky et al. \(2020\)](#) and [Black et al. \(2020\)](#). The results are presented in Table [A12](#) and, irrespective of the exact outcome, point to the same conclusions as our preferred specifications in Section [5](#). This strongly suggests that our results are not driven by sample selection related to who we can observe with valid outcomes, and thus strengthens the case for the external validity of the main estimates.

## 6 Robustness

In Section [5](#) we documented the causal effects of poor childhood health on long-run occupational success, and highlighted school attendance and disability as the intermediate outcomes and likely mechanisms. While we think of our estimates as capturing the effects of health deficiencies during childhood, we showed that poor childhood health could lead to disabilities later in life and thus our estimates partially encompass poor health that persists into adulthood and may impede employment opportunities or productivity. We also discussed the external validity of the results as well as selection into hospitalization and the linked sample. In this section, 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](#) to [6](#) for long-run outcomes, with the relative mobility measures in panels A and B, the occupational outcomes

in panels C to E, and where applicable, the occupational wages in panel F (Tables 4 and 5). In each table, we present our baseline estimates in the first column and show results using the hospitalized patient treatment indicator. In the online appendix, we reproduce the same specifications using the health deficiency index (Tables A13 to A15), and present similar robustness analyses for school attendance (Table A16) and disability (Tables A17 to A19).<sup>35</sup>

We begin the robustness analysis by considering factors that could generate a downward bias in our main estimates and hinder the interpretation of magnitudes presented in Section 5. For example, sub-groups characterized by high mortality may be positively selected in the linked, longitudinal samples, resulting in downward bias to our estimates. Furthermore, the same concern applies 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 may not necessarily lead to long-run consequences and treating these individuals as having poor health could likewise bias our estimates downward. 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, which suggests that our main estimates are not biased due to positive selection of surviving patients, the inclusion of mild health deficiencies with no long-term consequences of the treatment, and within-household contagion. Ruling out that our results are solely driven by poor health during infancy, when mortality was very high and where much prior research has focused, also supports our interpretation that, broadly speaking, poor health during childhood has negative long-run consequences.

An additional way in which we address potential contagion or unobserved hospitalization of siblings more directly is with the imputation of treatment status for control siblings. Figure A13 presents results where we impute the hospitalization indicator for 10 percent of randomly selected control siblings and repeat this exercise 1000 times. If our results were driven by either of these two phenomena we would expect the results to deviate from those presented in Section 5. Across

---

<sup>35</sup>When analyzing schooling and disability robustness we only present results for samples that pool observations across genders for the sake of brevity. Although this sample is not the preferred one for studying pre-existing disability, due to the relatively large effect sizes for females documented in Table 3, we have also verified that for the long-run sample (i.e. with only males) the robustness results all yield small estimates that are comparable to those from column 4 of Table 3.

all outcomes, the two sets of estimates are very similar, suggesting that these confounders are unlikely to be driving our results. Therefore, we conclude that these potential data limitations do not explain our main results.

The next set of robustness checks addresses different sources of within-household sample selection bias. 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 and find that our results are unchanged. 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. Thus, in column 3 we add families with multiple hospital patients to demonstrate that our results are unaffected by this sample restriction either.

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 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 individuals may be selectively linked. However, column 4 of Table 5 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.<sup>36</sup> Therefore, 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 and selection bias concerns due the set of siblings, households, residential locations, or hospitals included in the main analysis sample.

---

<sup>36</sup>Inpatient hospital admissions in nineteenth-century London were categorized as physician or surgeon patients. Among cohorts in our samples, physician patients accounted for 35 and 41 percent of inpatients at Barts and GOSH, respectively.

A related concern is within-household selection due to parents' differential sending of their children to the hospital based on unobserved health status. Columns 1 to 4 of Table A20 explore how sibling-specific factors affect the likelihood of hospitalization. Here, 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.

Next, we address the possibility that our results are driven by the selected sample of patients who can be followed across censuses. A potential source of sample selection bias is differential rates of linkage from the hospital records to the childhood censuses by health status. Table A21 examines how the value of a patient's health deficiency index 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, while column 4 considers the impact on linkage to any census using a pooled sample. We find some 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, however, 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 A14 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, and it is primarily patients admitted for conditions with a very high value of the health deficiency index (i.e. high in-hospital mortality) who are naturally less likely to be matched. While differences in linkage rates by health status of patients are small, they point, if anything, in the direction of a slight downward bias to our sibling fixed effects estimates. Lastly, we consider sample selection due to differential linkage rates between patients and siblings. We compare match rates between censuses in childhood and adulthood for patients and their siblings and find that patients are 1.8 pp or 2.7 percent more more likely to be linked, which is small in magnitude compared to differences in match rates by age or birth order. A plausible factor making patients more likely to be matched than their siblings is that they have already been matched from hospital records to a census during childhood. To the extent that

patients are positively selected on unobserved factors affecting the likelihood of a match, our estimates would likely be downward biased to a small degree. Overall, we conclude that sample selection concerns related to the health status of patients are unlikely to impart substantial bias our main results or qualitatively change our conclusions.

In addition, we assess whether our results are robust to modifications to our linking algorithm. 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. 2020). 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 more 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. (2020) who have shown in other contexts that coefficients of interest are typically stable across specifications with stricter versus more lenient match criteria. While we lack the “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 when allowing linked records to differ in first names or surnames. In Figure 1 we plot the coefficient on the patient indicator across the ten 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 precision decreases and our estimates shrink somewhat for some outcomes when we restrict to exactly matched names. The decline in magnitude applies exclusively to the results for the white collar or skilled occupational outcome while the results for the remaining labor market outcomes as well as all mechanisms are quantitatively similar even after restricting to exactly matched names.

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 neighboring birth cohorts of the outcome-year census is 20, where a similar name is defined as differing in the Jaro-Winkler distance score by less than 0.10. In Figure 2 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 0.10.<sup>37</sup> Our estimates are comparable across the various permutations though sometimes, as expected, we lose precision in samples with few similar names due to much reduced sample sizes.

Another potential concern relates to the fact that we do not observe the exact date of birth for children, and thus we use inferred age in our matching algorithm. More specifically, in the main specification, we allow for a mismatch of up to three years when linking between hospital and census records, but this tolerance for reporting errors could likewise result in false positive matches and bias our results. Thus, in Figure 3 we vary the restriction on the maximum age gap between linked sources from 0 to 3 years. Even when we consider the restriction that ages must match exactly in hospital-to-census linkages our results are substantively unchanged. Therefore, we discard this matching issue as a potential source of bias in our estimates.

Returning to Table 5, we present two additional robustness checks 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, due to the absence of birth place 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 to the same extent, 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. Overall, we conclude that our main results are not influenced by features of our approach to linking census data.

The final robustness discussion focuses on issues with inferring occupational status from occupational titles. A potential concern with our occupational mobility estimates is that they are

---

<sup>37</sup>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. We also run this analyses restricting similar names to be defined as having a Jaro-Winkler distance of less than 0.05, and the results remain unchanged.

based on a comparison between the occupational rank of father and the ranks of his sons with each individual's status observed at a single point in time. The literature has long recognized the issue of measurement error in inferring socioeconomic status from a single observation of an occupation (Solon 1992) and recent work has sought to address this by incorporating additional information on grandfather's status (Lindahl et al. 2015; Long and Ferrie 2018), dynastic status (Clark and Cummins 2020), uncles (Adermon et al. 2019), or by obtaining additional observations of occupational status for a given individual through multiple census linkages (Ward 2020).

While we lack additional information on the occupational status of patients and siblings, we address concerns about measurement error in father's status by incorporating two additional sources of information on socioeconomic status during childhood. First, one way in which the father's occupation may not accurately represent economic conditions of childhood household is that other breadwinners may co-reside in the same dwelling. For example, if the father is observed in the census at a young age, the primary breadwinner may be the children's grandfather or the father's older brother. Furthermore, mother's employment and occupation may also provide information on the household's status. Thus, we consider a robustness exercise in which we use the highest occupational rank among the father, the mother, and the household head (if the latter is someone other than one of the child's parents) for each household in the sample. Another benefit of inferring status from other household members is that our sample size increases as we bring in households for which the father's occupation is missing. Second, we also consider a slightly more conservative approach in which we only infer childhood status from other household members when the father's occupation was missing. Columns 2 and 3 of Table 6 report the results from these two estimation samples, respectively, and we find that the estimates are very similar to those obtained from the preferred specification, which suggests that measurement error in father's status is unlikely to be an important concern in our context.

An alternative approach to missing information on the sons and the father's occupational standing, which could generate biased estimates if these are not missing at random, is to provide bounds on our estimates (Lee 2009). Following this approach, we compare best versus worst case scenarios and assign the highest and lowest possible occupational ranks to individuals with missing information on occupational status. Columns 4 and 5 of Table 6 present these results for which our sample size increases by over 8 percent. With the exception of the indicator for upward mobility in the specification where we impute the lowest occupational class, all estimates remain statistically significant at conventional levels and point to the same qualitative conclusions. This exercise, thus, strengthens the notion that selection into who is observed with complete information on occupations is not an important driver of our main results.

Overall, the robustness exercises demonstrate that our main estimates for the effect of poor



childhood health on school attendance, disability, and occupational success 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, disability, 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 15.9 percent less likely to work in white collar occupations and 21.5 percent more likely to work in unskilled occupations as adults. In the first case, the effect size offsets 17.3 percent of the advantage of having a father in a white collar occupation. Consistent with these results, we also find that these patients were 10.6 percent more likely than their brothers to experience downward occupational mobility compared to the social standing of their father. Finally, we show that their occupational wages were 3.1 percent lower. Our results are robust to a variety of specification checks including issues related to selective mortality, pre-existing health deficiencies, sample selection choices, or changes to the matching algorithms. These results indicate that the persistence of poverty and within-household differences in occupational success can be attributed in part to inequality in childhood health.

In considering explanations for the estimated effects, we show that hospitalized boys were 2.9 percent less likely to attend school compared to their non-hospitalized brothers up to 10 years after the hospital admission. They were also more likely to suffer from severe disabilities as recorded in censuses both in childhood and adulthood. Although we cannot observe long-run outcomes for girls due to name changes at marriage, we find that the effects of health capital on both school

attendance and disability extend to girls. This implies that as adults these women might have also experienced detrimental socioeconomic outcomes.

The substantial heterogeneity in effects by cause of hospital admission suggests that the long-run consequences of poor childhood health can be attributed to not only the infectious fevers of childhood or chronic tuberculosis, which have been the focus of prior literature, but also to health conditions that were typically non-fatal and commonly occurring. Past research has overlooked these common illnesses given that our measurement of the historical health environment is mainly informed by mortality statistics. For example, our findings suggest that conditions that leave visible scars, including skin conditions such as eczema or congenital conditions such as cleft lip, could hurt an individual's occupational attainment in the long run.

Today, the rapid urbanization as well as changes in the disease environment and mortality trends taking place in developing countries are arguably similar to the historical experience of late-nineteenth and early-twentieth 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 countries, 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.

## Acknowledgements

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, Tymon Słoczyński, 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. We are also appreciative of the UK Data Archive and the Integrated Census Microdata (I-CeM) project at the University of Essex for providing access to the complete count census records for Great Britain. All errors of omission are our own.

## **Funding**

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.

## 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*, Vol. 102, No. 5, pp. 1832–1856.
- Abramitzky, Ran, Leah Platt Boustan, Katherine Eriksson, James J. Feigenbaum, and Santiago Pérez (2020) “Automated Linking of Historical Data,” *Journal of Economic Literature*, Vol. forthcoming.
- Abramitzky, Ran, Adeline Delavande, and Luis Vasconcelos (2011) “Marrying Up: The Role of Sex Ratio in Assortative Matching,” *American Economic Journal: Applied Economics*, Vol. 3, No. 3, pp. 124–157.
- Adermon, Adrian, Mikael Lindahl, and Marten Palme (2019) “Dynastic Human Capital and Intergenerational Mobility,” Working Paper.
- Adhvaryu, Achyuta, Steven Bednar, Teresa Molina, Quynh Nguyen, and Anant Nyshadham (2020) “When It Rains It Pours: The Long-run Economic Impacts of Salt Iodization in the United States,” *Review of Economics and Statistics*, Vol. 102, No. 2, pp. 395–407.
- Almond, Douglas (2006) “Is the 1918 Influenza Pandemic Over? Long-Term Effects of In Utero Influenza Exposure in the Post-1940 U.S. Population,” *Journal of Political Economy*, Vol. 114, No. 4, pp. 672–712.
- Almond, Douglas and Janet Currie (2011a) “Human Capital Development Before Age Five,” in Orley Ashenfelter and David Card eds. *Handbook of Labor Economics*, Vol. 4B: Amsterdam: Elsevier, pp. 1315–1486.
- (2011b) “Killing Me Softly: The Fetal Origins Hypothesis,” *Journal of Economic Perspectives*, Vol. 25, No. 3, pp. 153–172.
- Almond, Douglas, Janet Currie, and Valentina Duque (2018) “Childhood Circumstances and Adult Outcomes: Act II,” *Journal of Economic Literature*, Vol. 56, No. 4, pp. 1360–1446.
- Armstrong, W.A. (1972) “The Use of Information About Occupation,” in E. A. Wrigley ed. *Nineteenth-Century Society: Essays in the Use of Quantitative Methods for the Study of Social Data*, Cambridge: Cambridge University Press, pp. 191–310.
- Arthi, Vellore and Eric B. Schneider (2020) “Infant Feeding and Post-Weaning Health: Evidence from Turn-of-the-Century London,” Working Paper.
- Bailey, Martha, Connor Cole, Morgan Henderson, and Catherine Massey (2020) “How Well Do Automated Methods Linking Perform? Evidence from U.S. Historical Data,” *Journal of Economic Literature*, Vol. forthcoming.
- 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*, Vol. 131, No. 4, pp. 1637–1680.

- Barreca, Alan I. (2010) “The Long-Term Economic Impact of in Utero and Postnatal Exposure to Malaria,” *Journal of Human Resources*, Vol. 45, No. 4, pp. 865–892.
- 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*, Vol. No. 24725.
- Beach, Brian, Joseph Ferrie, Martin Saavedra, and Werner Troesken (2016) “Typhoid Fever, Water Quality, and Human Capital Formation,” *Journal of Economic History*, Vol. 76, No. 1, pp. 41–75.
- 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.
- Black, Sandra, Sanni Breining, David Figlio, Jonathan Guryan, Krzysztof Karbownik, Helena Skyt Nielsen, Jeffrey Roth, and Marianne Simonsen (2020) “Sibling Spillovers,” *Economic Journal*, Vol. forthcoming.
- Bleakley, Hoyt (2007) “Disease and Development: Evidence from Hookworm Eradication in the American South,” *Quarterly Journal of Economics*, Vol. 122, No. 1, pp. 73–117.
- (2010) “Malaria Eradication in the Americas: A Retrospective Analysis of Childhood Exposure,” *American Economic Journal: Applied Economics*, Vol. 2, No. 2, pp. 1–45.
- Brown, David W, Amanda E Kowalski, and Ithai Z Lurie (2020) “Long-Term Impacts of Childhood Medicaid Expansions on Outcomes in Adulthood,” *Review of Economic Studies*, Vol. 87, No. 2, pp. 792–821.
- 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,” *Review of Economics and Statistics*, Vol. 101, No. 2, pp. 1–14.
- Bütikofer, Aline and Kjell G. Salvanes (2020) “Disease Control and Inequality Reduction: Evidence from a Tuberculosis Testing and Vaccination Campaign,” *Review of Economic Studies*, Vol. 87, No. 5, pp. 2087–2125.
- Carpenter, Mary Wilson (2010) *Health, Medicine, and Society in Victorian England*, Santa Barbara, CA: Praeger.
- Case, Anne, Angela Fertig, and Christina Paxson (2005) “The Lasting Impact of Childhood Health and Circumstance,” *Journal of Health Economics*, Vol. 24, No. 2, pp. 365 – 389.
- Case, Anne, Darren Lubotsky, and Christina Paxson (2002) “Economic Status and Health in Childhood: The Origins of the Gradient,” *American Economic Review*, Vol. 92, No. 5, pp. 1308–1334.
- Case, Anne and Christina Paxson (2008) “Stature and Status: Height, Ability, and Labor Market Outcomes,” *Journal of Political Economy*, Vol. 116, No. 3, pp. 499–532.

- (2010) “Causes and Consequences of Early-Life Health,” *Demography*, Vol. 47, No. 1, pp. S65–S85.
- Chatto and Windus (1897) “Sixty Years of Healing,” *Belgravia: A London Magazine*, Vol. 93, pp. 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, pp.xii, 364 pages.
- Clark, Gregory and Neil Cummins (2020) “Family Matters? Do Relatives Other Than Parents Matter to Social Outcomes, England 1780-2018?,” Working Paper.
- Cohen, Deborah (2014) *Family Secrets: The Things we Tried to Hide*, London: Penguin Books.
- Costa, Dora L. (2015) “Health and the Economy in the United States from 1750 to the Present,” *Journal of Economic Literature*, Vol. 53, No. 3, pp. 503–570.
- Cullis, P. S. and D. G. Young (2013) “Nineteenth Century Mortality at the Royal Hospital for Sick Children, Glasgow,” *Scottish Medical Journal*, Vol. 58, No. 3, pp. 182–187.
- Currie, Janet (2009) “Healthy, Wealthy, and Wise: Socioeconomic Status, Poor Health in Childhood, and Human Capital Development,” *Journal of Economic Literature*, Vol. 47, No. 1, pp. 87–122.
- Currie, Janet, Mark Stabile, Phongsack Manivong, and Leslie L. Roos (2010) “Child Health and Young Adult Outcomes,” *Journal of Human Resources*, Vol. 45, No. 3, pp. 517–548.
- Currie, Janet and Tom Vogl (2013) “Early-Life Health and Adult Circumstance in Developing Countries,” *Annual Review of Economics*, Vol. 5, No. 1, pp. 1–36.
- Cutler, David, Angus Deaton, and Adriana Lleras-Muney (2006) “The Determinants of Mortality,” *Journal of Economic Perspectives*, Vol. 20, No. 3, pp. 97–120.
- Dahis, Ricardo, Emily Nix, and Nancy Qian (2020) “Choosing Racial Identity in the United States, 1880-1940,” Working Paper.
- Doyle, Shane, Felix Meier zu Selhausen, and Jacob Weisdorf (2020) “The Blessings of Medicine? Patient Characteristics and Health Outcomes in a Ugandan Mission Hospital, 1908-1970,” *Social History of Medicine*, Vol. 33, No. 3, pp. 946–980.
- Feigenbaum, James (2018) “Multiple Measures of Historical Intergenerational Mobility: Iowa 1915 to 1940,” *Economic Journal*, Vol. 128, No. 612, pp. F446–F481.
- Feigenbaum, James J. (2016) “Automated Census Record Linking: A Machine Learning Approach,” Working Paper.



- Feinstein, Charles (1988) “The Rise and Fall of the Williamson Curve,” *The Journal of Economic History*, Vol. 48, No. 3, p. 699–729.
- 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*, Vol. 29, No. 4, pp. 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, pp.1 v.
- Fogel, Robert William (2004) *The Escape from Hunger and Premature Death, 1700-2100*, Cambridge: Cambridge University Press, pp.xx, 191 p.
- Franklin, Alfred White (1964) “Children’s Hospitals,” in F. N. L. Poynter ed. *The Evolution of Hospitals in Britain*, London: Pitman, pp. 103–121.
- Gensowski, Miriam, Torben Heien Nielsen, Nete Munk Nielsen, Maya Rossin-Slater, and Miriam Wüst (2019) “Childhood Health Shocks, Comparative Advantage, and Long-Term Outcomes: Evidence from the Last Danish Polio Epidemic,” *Journal of Health Economics*, Vol. 66, pp. 27–36.
- Gertler, Paul and Jonathan Gruber (2002) “Insuring Consumption against Illness,” *American Economic Review*, Vol. 92, No. 1, pp. 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,” *Review of Economic Studies*, Vol. 78, No. 3, pp. 938–973.
- Hardy, Anne (2001) *Health and Medicine in Britain since 1860*, New York: Palgrave, pp.xi, 220 p.
- Hawkins, Sue and Andrea Tanner (2013) “Myth, Marketing, and Medicine: Life in British Children’s Hospitals, 1850-1914,” in Laurinda Abreu and Sally Sheard eds. *Hospital Life: Theory and Practice from the Medieval to the Modern*, Bern, Switzerland: Peter Lang, pp. 209–236.
- Heckman, James (2006) “Skill Formation and the Economics of Investing in Disadvantaged Children,” *Science*, Vol. 312, No. 5782, pp. 1900–1902.
- Higgs, Michelle (2009) *Life in the Victorian Hospital*, Stroud, Gloucestershire: History Press, pp.160.
- 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*, Vol. 9, No. 4, pp. 78–104.
- Hoehn-Velasco, Lauren (2020) “The Long-term Impact of Preventative Public Health Programs,” *Economic Journal*, Vol. forthcoming.

- Hong, Sok Chul (2007) “The Burden of Early Exposure to Malaria in the United States, 1850-1860: Malnutrition and Immune Disorders,” *Journal of Economic History*, Vol. 67, No. 4, pp. 1001–1035.
- (2013) “Malaria: An Early Indicator of Later Disease and Work Level,” *Journal of Health Economics*, Vol. 32, No. 3, pp. 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*, Vol. 38, No. 3, pp. 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*, Vol. 2, No. 2, pp. 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 (2020) “Infant Health and Later-Life Labor Market Outcomes: Evidence from the Introduction of Sulpha Antibiotics in Sweden,” *Journal of Human Resources*, Vol. forthcoming.
- Lee, David (2009) “Training, Wages, and Sample Selection: Estimating Sharp Bounds on Treatment Effects,” *Review of Economic Studies*, Vol. 76, No. 3, pp. 1071–1102.
- van Leeuwen, Marco H. D. and Ineke Maas (2011) *Hisclass: A Historical International Social Class Scheme*, Leuven, Belgium: Leuven University Press, pp.181.
- van Leeuwen, Marco H. D., Ineke Maas, and Andrew Miles (2002) *Hisco: Historical International Standard Classification of Occupations*, Leuven, Belgium: Leuven University Press, pp.441.
- Lindahl, Mikael, Marten Palme, Sofia Sandgren Massih, and Anna Sjogren (2015) “Long-term Intergenerational Persistence of Human Capital: An Empirical Analysis of Four Generations,” *Journal of Human Resources*, Vol. 50, No. 1, pp. 1–33.
- 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, pp.vii, 217.
- Long, Jason (2006) “The Socioeconomic Return to Primary Schooling in Victorian England,” *The Journal of Economic History*, Vol. 66, No. 4, p. 1026–1053.
- (2013) “The Surprising Social Mobility of Victorian Britain,” *European Review of Economic History*, Vol. 17, No. 1, pp. 1–23.
- Long, Jason and Joseph Ferrie (2013) “Intergenerational Occupational Mobility in Great Britain and the United States since 1850,” *American Economic Review*, Vol. 103, No. 4, pp. 1109–1137.
- (2018) “Grandfathers Mattered(ed): Occupational Mobility Across Three Generations in the US and Britain, 1850-1911,” *Economic Journal*, Vol. 128, No. 612, pp. F422–F445.

- McKeown, Thomas (1976) *The Modern Rise of Population*, London: Edward Arnold, pp.[5], 168 pages.
- Mercer, Alexander (2014) *Infections, Chronic Disease, and the Epidemiological Transition : A New Perspective*, Rochester, NY: University of Rochester Press, pp.338.
- 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*, Vol. No. 26174.
- Milner, Benjamin (2020) “The Impact of State-Provided Education: Evidence from the 1870 Education Act,” Working Paper.
- Mokyr, Joel (2000) “Why “More Work for Mother?” Knowledge and Household Behavior, 1870-1945,” *Journal of Economic History*, Vol. 60, No. 1, pp. 1–41.
- 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*, Vol. 105, No. 8, pp. 1–31.
- Parman, John (2015a) “Childhood Health and Sibling Outcomes: Nurture Reinforcing Nature During the 1918 Influenza Pandemic,” *Explorations in Economic History*, Vol. 58, pp. 22–43.
- (2015b) “Childhood health and human capital: New evidence from genetic brothers in arms,” *Journal of Economic History*, Vol. 75, No. 1, pp. 30–64.
- Pérez, Santiago (2019) “Intergenerational Occupational Mobility across Three Continents,” *Journal of Economic History*, Vol. 79, No. 2, pp. 383–416.
- Ruggles, Steven (2014) “Big Microdata for Population Research,” *Demography*, Vol. 51, No. 1, pp. 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*, Vol. 64, pp. 62–81.
- Schürer, K and Joe Day (2019) “Migration to London and the Development of the North–South Divide, 1851–1911,” *Social History*, Vol. 44, No. 1, pp. 26–56.
- Schwandt, Hannes (2018) “The Lasting Legacy of Seasonal Influenza: In-Utero Exposure and Labor Market Outcomes,” Working Paper.
- Słoczyński, Tymon (2020) “Interpreting OLS Estimands When Treatment Effects are Heterogeneous: Smaller Groups Get Larger Weights,” *Review of Economics and Statistics*, Vol. forthcoming.
- Smith, James P. (2009) “The Impact of Childhood Health on Adult Labor Market Outcomes,” *Review of Economics and Statistics*, Vol. 91, No. 3, pp. 478–489.

- Solon, Gary (1992) “Intergenerational Income Mobility in the United States,” *The American Economic Review*, Vol. 82, No. 3, pp. 393–408.
- 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.
- Venkataramani, Atheendar S. (2012) “Early Life Exposure to Malaria and Cognition in Adulthood: Evidence from Mexico,” *Journal of Health Economics*, Vol. 31, No. 5, pp. 767–780.
- Wang, Haidong, Kaja M Abbas, Mitra Abbasifard, Mohsen Abbasi-Kangevari, Hedayat Abastabar, Foad Abd-Allah, Ahmed Abdelalim, Hassan Abolhassani, Lucas Guimarães Abreu, Michael R M Abrigo, Abdelrahman I Abushouk, Maryam Adabi, Tim Adair, Oladimeji M Adedayo, Isaac Akinkunmi Adedeji, Victor Adekanmbi, Abiodun Moshood Adeoye, Olatunji O Adetokunboh, Shailesh M Advani, ..., and Christopher J L Murray (2020) “Global age-sex-specific fertility, mortality, healthy life expectancy (HALE), and population estimates in 204 countries and territories, 1950-2019: a comprehensive demographic analysis for the Global Burden of Disease Study 2019,” *The Lancet*, Vol. 396, No. 10258, pp. 1160–1203.
- Ward, Zach (2020) “Intergenerational Mobility in American History: Accounting for Race and Measurement Error,” Working Paper.
- Warren, John, Laurie Knies, Steven Haas, and Elaine Hernandez (2012) “The Impact of Childhood Sickness on Adult Socioeconomic Outcomes: Evidence from Late 19th Century America,” *Social Science and Medicine*, Vol. 75, No. 8, pp. 1531–1538.
- Williamson, Jeffrey G. (1980) “Earnings Inequality in Nineteenth-Century Britain,” *The Journal of Economic History*, Vol. 40, No. 3, pp. 457–475.
- (1982) “The Structure of Pay in Britain, 1710–1911,” *Research in Economic History*, Vol. 7, pp. 1–54.
- Wohl, Anthony S. (1983) *Endangered Lives: Public Health in Victorian Britain*, Cambridge, MA: Harvard University Press, pp.440 p.
- Worboys, Michael (2000) *Spreading Germs: Diseases, Theories, and Medical Practice in Britain, 1865-1900*, New York: Cambridge University Press, pp.xvi, 327 p.
- 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 (0.064)	1,243 (0.054)	1,777 (0.167)	983 (0.035)
Multiple matches	10,603 (0.667)	15,318 (0.663)	6,255 (0.586)	17,702 (0.622)
Unique match	4,284 (0.269)	6,560 (0.284)	2,635 (0.247)	9,761 (0.343)
Sibling present	4,007 (0.252)	5,938 (0.257)	2,240 (0.210)	9,167 (0.322)
Long-run censuses	1901, 1911	1901, 1911	1911	1901, 1911
Panel B: Linkage to census in adulthood				
Patient matched	2,968 (0.187)	3,747 (0.162)	1,380 (0.129)	6,236 (0.219)
Patient and sibling	1,194 (0.075)	1,565 (0.068)	362 (0.034)	3,097 (0.109)
In final sample	900 (0.057)	1,008 (0.044)	239 (0.022)	2,146 (0.075)
Total admissions	15,903	23,121	10,667	28,446

Notes: Panel A presents sample sizes (and linkage rates in parentheses) from matching inpatient hospital admission records to the 1881, 1891, and 1901 censuses (columns 1 to 3), as well from pooling 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 no match, multiple matches, 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 shows the number of admissions by 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. The 2,146 admissions in the final sample in column 4 represent the total number of admissions by the 1,849 patients (and households) in the main estimation sample in Table 2 since some patients were admitted to the hospital more than once.

Table 2: Intergenerational mobility, occupational status, and log occupational wages

	Mobility		Own rank			(6) Log wage
	(1) Class ↗	(2) Class ↘	(3) White collar	(4) Skilled +	(5) Unskilled	
Panel A: Effects of hospital admission						
Patient	-0.030** (0.014)	0.028** (0.013)	-0.043*** (0.014)	-0.036** (0.016)	0.033*** (0.012)	-0.032*** (0.012)
% effect	8.5	10.6	15.9	6.7	21.5	3.1
Panel B: Effects of health deficiency index						
Health deficiency index	-0.098** (0.046)	0.084* (0.043)	-0.142*** (0.046)	-0.116** (0.054)	0.103*** (0.039)	-0.096** (0.038)
% effect ( $\sigma$ )	3.5	4.1	6.7	2.8	8.7	1.2
Mean of Y	0.357	0.260	0.273	0.539	0.153	4.628
N Households	1,849	1,849	1,849	1,849	1,849	1,849
N	3,698	3,698	3,698	3,698	3,698	3,698

Notes: The dependent variables in columns 1 and 2 are measures of intergenerational occupational mobility: the probability of having a higher or lower occupational status than one's father, respectively. In columns 3 to 5, the dependent variables are indicators of own occupational rank: the probabilities of entering a white-collar occupation; a white-collar or skilled occupation; or an unskilled occupation. In column 6, the dependent variable is an individual's log occupational wage. Each regression includes sibling fixed effects as well as age-by-outcome census year and birth order fixed effects, an indicator for the older sibling in each pair of observations from a household, fixed effects for the number of children age 0 to 5 in the household during one's infancy (age 0 to 2) and at the age of the first hospital admission among siblings in the household, a standardized measure 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). Standard errors are clustered by childhood household. Percent effects for log wages (column 6) are calculated using the formula  $100 \times \exp(\beta) - 1$ . Percent effects in panel B (columns 1 to 5) are computed as  $\beta \times \sigma_{\text{HDI}}$  scaled by the dependent variable mean, where  $\sigma_{\text{HDI}} = 0.13$  is the standard deviation of the health deficiency index in the hospital population.

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

Table 3: Mechanisms

	Any disability x 100									
	Schooling			Pre-existing disability			Childhood disability			LR disability
	(1) Males	(2) Females	(3) Both sexes	(4) Males	(5) Females	(6) Both sexes	(7) Males	(8) Females	(9) Both sexes	(10) Males
Panel A: Effects of hospital admission										
Patient	-0.021*	-0.032**	-0.026***	-0.025	0.221	0.101	0.539**	0.925***	0.746***	0.739***
	(0.011)	(0.013)	(0.008)	(0.190)	(0.151)	(0.118)	(0.234)	(0.288)	(0.185)	(0.233)
% effect	2.9	4.3	3.5	16.4	141.7	65.8	118.5	161.7	145.6	127.5
Panel B: Effects of health deficiency index										
Health deficiency index	-0.079**	-0.081**	-0.079***	-0.148	0.150	0.029	1.841**	2.516***	2.235***	1.866**
	(0.039)	(0.041)	(0.028)	(0.467)	(0.453)	(0.303)	(0.745)	(0.875)	(0.569)	(0.799)
% effect ( $\sigma$ )	1.4	1.4	1.4	12.6	12.4	2.4	52.2	56.7	56.3	41.5
Mean of Y	0.750	0.736	0.743	0.152	0.156	0.154	0.455	0.572	0.513	0.580
N	1,530	1,510	3,040	4,608	4,488	9,096	5,718	5,594	11,312	4,312

Notes: In columns 1 to 3, the dependent variable is an indicator for school attendance (see section 3.6 for details on how the variable is coded). In each schooling-outcome regression (columns 1 to 3), the sample includes children aged 5 to 10 at the time of census enumeration. In the remaining columns, the dependent variable is an indicator variable, that is equal to one if an individual is recorded as having a disability in the census, multiplied by 100. In columns 4 to 6 the outcome is pre-existing disability, which is reported in the census up to 10 years *before* hospital admission. Columns 7 to 9 present estimates for childhood disability, which is recorded in the census up to 10 years *after* hospital admission. Column 10 shows estimates for disability in adulthood for boys only and involves linkages to two censuses. See section 3.7 for a description of how the variable is coded and for a discussion of the sample construction. In panel A, the explanatory variable of interest is an indicator for hospital admission, and in panel B, it is the health deficiency index. All regressions include sibling fixed effects as well as age-by-census year and birth-order fixed effects, an indicator for the older sibling in each pair of observations from a household, fixed effects for the number of children age 0 to 5 in the household during one's infancy (age 0 to 2) and at the age of the first hospital admission among siblings in the household, a standardized measure of first name frequency, an interaction of first name and surname frequency, and indicators for match quality (exact matches on first name and surname). For each group of specifications, the third column pools together the samples in the first two columns and makes same-gender comparisons within households. Standard errors are clustered by childhood household. Percent effects in panel B are computed as  $\beta \times \sigma_{\text{HDI}}$  scaled by the dependent variable mean, where  $\sigma_{\text{HDI}} = 0.13$  is the standard deviation of the health deficiency index in the hospital population.

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(Class ↗)						
Patient	-0.030** (0.014)	-0.034** (0.014)	-0.034** (0.015)	-0.027* (0.014)	-0.032** (0.015)	-0.038** (0.015)
Mean of Y	0.357	0.356	0.355	0.359	0.357	0.356
Panel B: Effects on P(Class ↘)						
Patient	0.028** (0.013)	0.027** (0.013)	0.029** (0.014)	0.023* (0.013)	0.026* (0.013)	0.026* (0.014)
Mean of Y	0.260	0.256	0.261	0.261	0.260	0.263
Panel C: Effects on P(White collar)						
Patient	-0.043*** (0.014)	-0.044*** (0.015)	-0.050*** (0.015)	-0.048*** (0.015)	-0.049*** (0.015)	-0.049*** (0.015)
Mean of Y	0.273	0.269	0.278	0.276	0.267	0.276
Panel D: Effects on P(Skilled +)						
Patient	-0.036** (0.016)	-0.040** (0.017)	-0.042** (0.017)	-0.030* (0.017)	-0.035** (0.017)	-0.037** (0.018)
Mean of Y	0.539	0.538	0.541	0.546	0.534	0.541
Panel E: Effects on P(Unskilled)						
Patient	0.033*** (0.012)	0.032*** (0.012)	0.030** (0.012)	0.031** (0.012)	0.033*** (0.012)	0.038*** (0.013)
Mean of Y	0.153	0.155	0.151	0.153	0.152	0.154
Panel F: Effects on log occupational wage						
Patient	-0.032*** (0.012)	-0.034*** (0.012)	-0.035*** (0.012)	-0.036*** (0.012)	-0.037*** (0.012)	-0.040*** (0.013)
Mean of Y	4.628	4.627	4.629	4.630	4.624	4.634
N	3,698	3,328	3,328	3,350	3,256	3,044

Notes: Each cell displays the coefficient for the hospitalization indicator from a separate regression. The dependent variables in panels A to F mirror those in columns 1 to 6 of Table 2. Column 1 reproduces the estimates in panel A of Table 2. Columns 2 and 3 drop households with patients admitted for conditions in the top and the bottom 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 the 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 hhllds.	(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(Class ↗)							
Patient	-0.030** (0.014)	-0.021* (0.013)	-0.020 (0.013)	-0.039** (0.015)	-0.028* (0.014)	-0.037** (0.014)	-0.046*** (0.017)
Mean of Y	0.357	0.355	0.355	0.367	0.357	0.353	0.363
Panel B: Effects on P(Class ↘)							
Patient	0.028** (0.013)	0.027** (0.012)	0.026** (0.012)	0.027** (0.014)	0.031** (0.014)	0.023* (0.013)	0.025 (0.016)
Mean of Y	0.260	0.259	0.260	0.265	0.261	0.258	0.267
Panel C: Effects on P(White collar)							
Patient	-0.043*** (0.014)	-0.038*** (0.013)	-0.037*** (0.013)	-0.046*** (0.015)	-0.043*** (0.015)	-0.047*** (0.015)	-0.053*** (0.017)
Mean of Y	0.273	0.275	0.274	0.278	0.279	0.276	0.286
Panel D: Effects on P(Skilled +)							
Patient	-0.036** (0.016)	-0.029* (0.015)	-0.027* (0.015)	-0.041** (0.018)	-0.043** (0.017)	-0.039** (0.017)	-0.047** (0.020)
Mean of Y	0.539	0.539	0.539	0.546	0.550	0.540	0.550
Panel E: Effects on P(Unskilled)							
Patient	0.033*** (0.012)	0.022** (0.011)	0.023** (0.011)	0.038*** (0.013)	0.031** (0.012)	0.030** (0.012)	0.026* (0.015)
Mean of Y	0.153	0.156	0.155	0.144	0.145	0.154	0.142
Panel F: Effects on log occupational wage							
Patient	-0.032*** (0.012)	-0.025** (0.011)	-0.024** (0.011)	-0.032** (0.013)	-0.034*** (0.012)	-0.036*** (0.012)	-0.043*** (0.015)
Mean of Y	4.628	4.629	4.629	4.639	4.634	4.628	4.650
N	3,698	4,473	4,512	3,032	3,332	3,302	2,366

Notes: Each cell displays the coefficient for the hospitalization indicator from a separate regression. The dependent variables in panels A to F mirror those in columns 1 to 6 of Table 2. 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 the 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 6: Long-run outcomes: Robustness to variations in occupational status

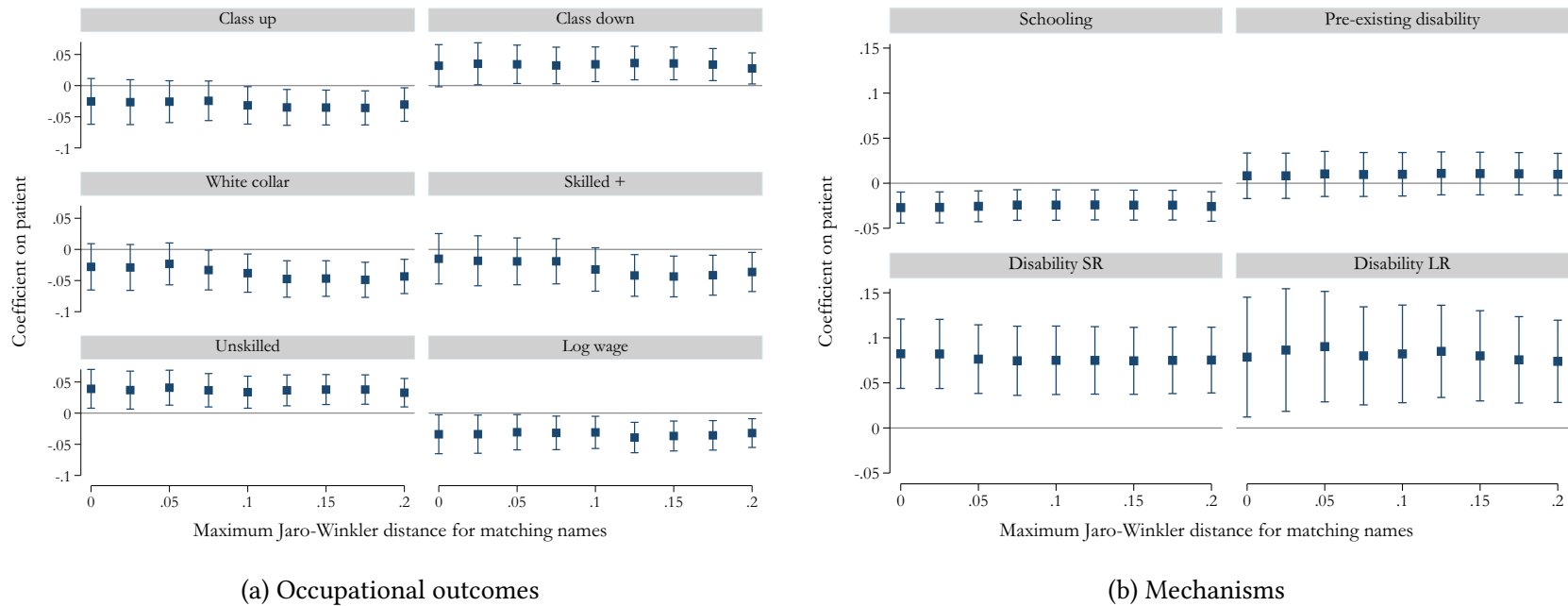
	(1) Baseline estimate	(2) Impute Hhld SES	(3) Highest Hhld SES	(4) High class if missing	(5) Low class if missing
Panel A: Effects on P(Class ↗)					
Patient	-0.030** (0.014)	-0.028** (0.013)	-0.032** (0.013)	-0.026* (0.013)	-0.017 (0.014)
Mean of Y	0.357	0.346	0.360	0.339	0.386
Panel B: Effects on P(Class ↘)					
Patient	0.028** (0.013)	0.028** (0.013)	0.028** (0.012)	0.030** (0.013)	0.021* (0.012)
Mean of Y	0.260	0.273	0.258	0.297	0.251
Panel C: Effects on P(White collar)					
Patient	-0.043*** (0.014)	-0.042*** (0.014)	-0.042*** (0.014)	-0.041*** (0.013)	-0.033** (0.014)
Mean of Y	0.273	0.270	0.270	0.259	0.306
Panel D: Effects on P(Skilled +)					
Patient	-0.036** (0.016)	-0.039** (0.016)	-0.039** (0.016)	-0.038** (0.016)	-0.029* (0.016)
Mean of Y	0.539	0.542	0.542	0.511	0.558
Panel E: Effects on P(Unskilled)					
Patient	0.033*** (0.012)	0.034*** (0.011)	0.034*** (0.011)	0.029** (0.011)	0.037*** (0.012)
Mean of Y	0.153	0.151	0.151	0.147	0.194
N	3,698	3,870	3,870	4,004	4,004

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. Column 1 reproduces the estimates in panel A of Table 2 which excludes observations with missing adult occupation or father's occupation. Column 2 uses the highest occupational rank among the father, mother, and the household head as a measure of childhood socioeconomic status (SES) while column 3 sequentially uses the mother's occupation or the occupation of the household head in place of the father's occupation in measuring childhood SES. Columns 4 and 5 impute the highest and lowest occupational class, respectively, for individuals with missing occupations. See Table 2 for a description of the 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.

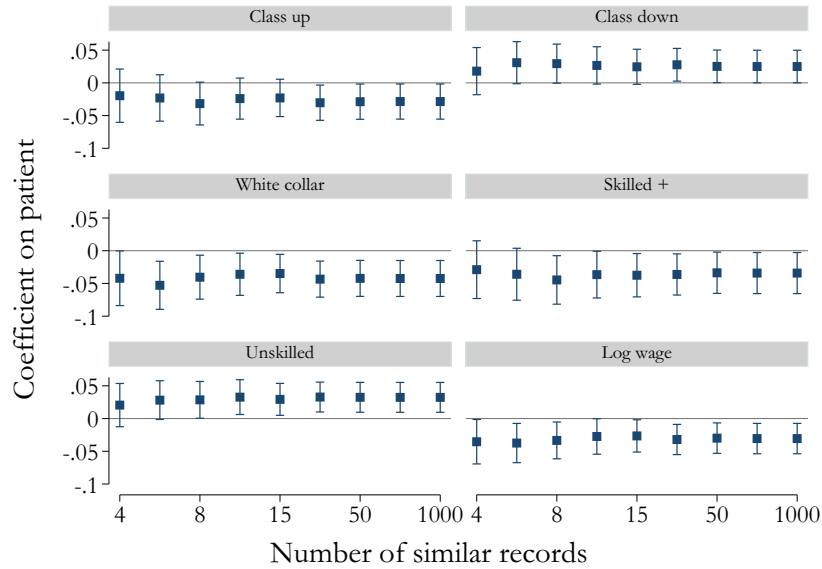
## 9 Figures

Figure 1: Robustness to changing Jaro-Winkler distance threshold

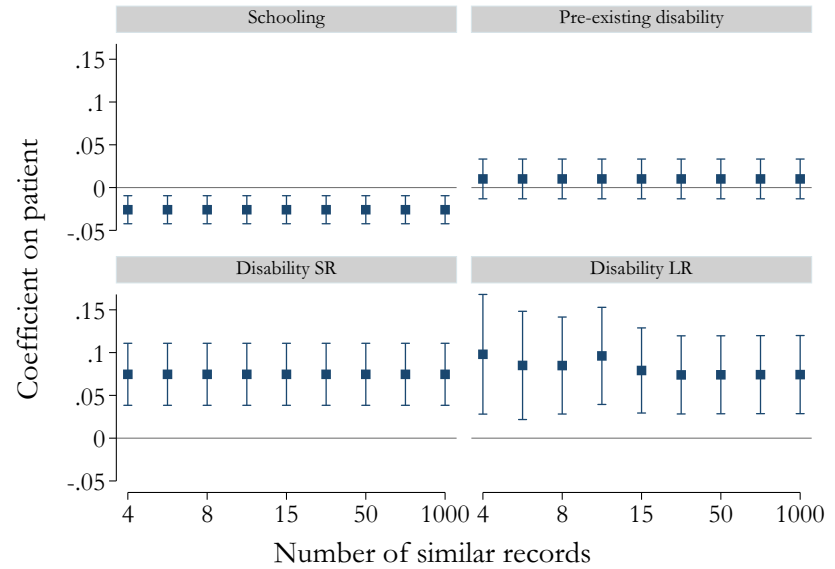


Notes: This figure presents estimated coefficients on the patient indicator variable and 95-percent confidence intervals from separate regressions in which we decrease the Jaro-Winkler distance threshold for inclusion in the sample, moving from the right to the left side of the figures. The long-run occupational outcomes are shown in figure 1a and the mechanisms are shown in figure 1b. The left side ( $x = 0$ ) of each figure 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. See tables 2 and 3 for a description of the empirical specifications in figures 1a and 1b, respectively. The disability outcomes are multiplied by 10 for ease of visualization. With the exception of the long-run disability outcome, which is based on a sample of linked males only, the disability and schooling outcomes use samples that are pooled across genders. Standard errors are clustered by childhood household.

Figure 2: Robustness to changing similar names threshold (JW = 0.10)



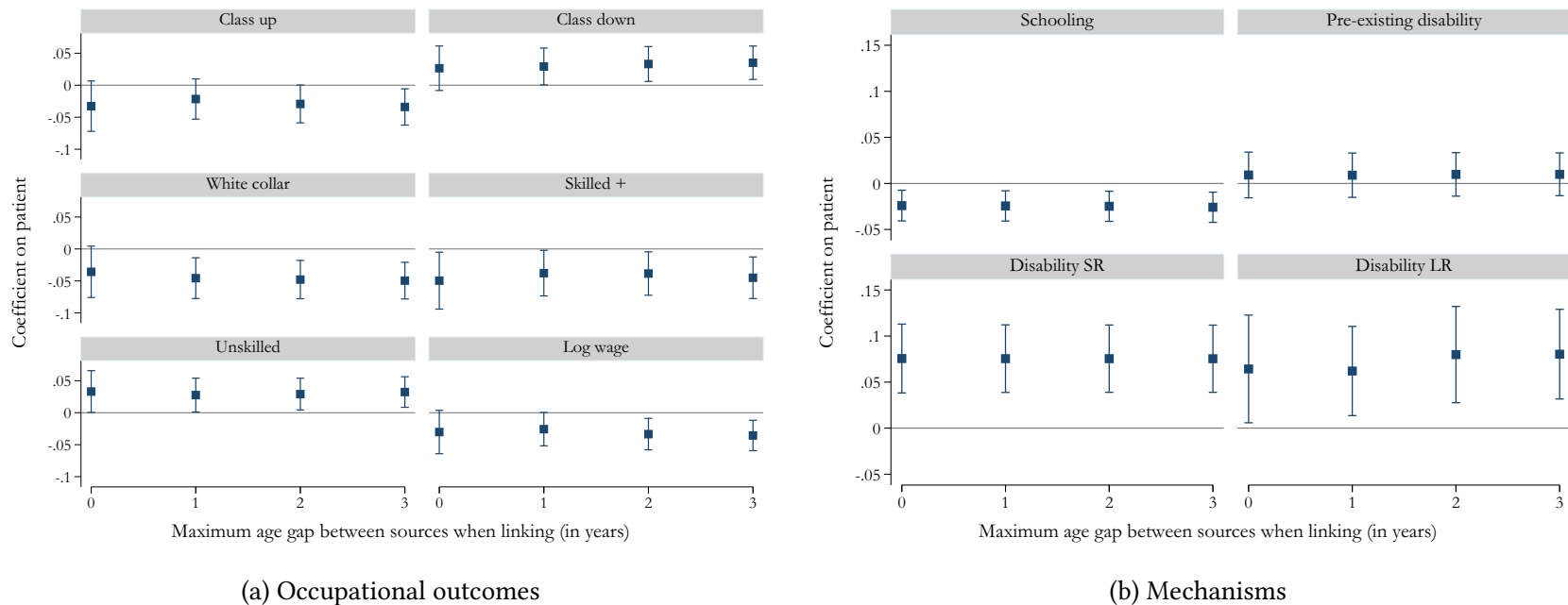
(a) Occupational outcomes



(b) Mechanisms

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 records permitted to have similar names that are within one year of birth 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. Similar names are defined as differing in Jaro-Winkler scores by less than 0.10 for the first and last names compared to the name of the matched individual. The long-run occupational outcomes are shown in figure 2a and the mechanisms are shown in figure 2b. The main specification corresponds to the estimates with 20 similar names and a Jaro-Winkler threshold of 0.10 for a name to be similar to the matched record. See tables 2 and 3 for a description of the empirical specifications in figures 2a and 2b, respectively. The disability outcomes are multiplied by 10 for ease of visualization. With the exception of the long-run disability outcome, which is based on a sample of linked males only, the disability and schooling outcomes use samples that are pooled across genders. Standard errors are clustered by childhood household.

Figure 3: Robustness to changing age-gap threshold for linking



Notes: This figure presents estimated coefficients on the patient indicator variable and 95-percent confidence intervals from separate regressions in which we decrease the threshold for differences in reported ages when linking between hospital and census records or between multiple censuses, moving from the right to the left side of the figures. The long-run occupational outcomes are shown in figure 3a and the mechanisms are shown in figure 3b. The left side ( $x = 0$ ) of each figure corresponds to the restriction that ages must match exactly across censuses and hospital-to-census linkages, while the right side ( $x = 3$ ) corresponds to the main sample where differences in ages of up to 3 years are tolerated. See tables 2 and 3 for a description of the empirical specifications in figures 3a and 3b, respectively. The disability outcomes are multiplied by 10 for ease of visualization. With the exception of the long-run disability outcome, which is based on a sample of linked males only, the disability and schooling outcomes use samples that are pooled across genders. Standard errors are clustered by childhood household.



# 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.<sup>1</sup> 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 \(2015a\)](#) 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.<sup>2</sup> 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](#)).

---

<sup>1</sup>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.

<sup>2</sup>See [Dahis et al. \(2020\)](#) and the supplemental materials to [Parman \(2015a\)](#) 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. \(2020\)](#).<sup>3</sup> 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 3 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.<sup>4</sup>
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  $(1 + 0.1)J_s \geq \bar{J}$ , 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.

---

<sup>3</sup>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. \(2020\)](#) that using a combination of samples generated with the [Ferrie \(1996\)](#) and [Feigenbaum \(2016\)](#) methods results in a much lower Type I error rate.

<sup>4</sup>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 [A22](#). The share of unique matches ranges from 51 to 64 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. As with the census-to-census linkages, we require the age to differ by no more than 3 years between sources. 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 [A23](#) 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, and 34 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 samples used in the main analysis of occupational outcomes and disability in adulthood are 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 school-aged individuals at the time of 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.

When we turn to the short-run disability outcomes, we follow the same procedure as for schooling, with the exceptions that we allow children to be enumerated in the census at ages 0 to 21 instead of restricting to ages 5 to 10, and we include links to the 1901 or 1911 censuses. For disability during childhood, we also require the hospital admission to have occurred prior to enumeration in the census, while the opposite holds true for the pre-existing disability sample.

#### **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 choose the older sibling if the patient's unique identifier in the I-CeM complete count files is an even number and the younger sibling if it is an odd number, 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.

## A.5 Linked sample of fathers

Thus far, we have described methods for linking population censuses over time and to hospital admission records, which we use to generate our main analysis sample. In an auxiliary analysis discussed in section 5.1 and reported in Figure A8, we also make use of a sample of fathers linked between adjacent censuses. The starting points for the linked sample of fathers are the samples of children age 0 to 21 linked from the 1881 to 1891 and 1891 to 1901 censuses. We append to these samples pairs of census records that are linked to the same hospital patients. We require children to be enumerated in the same household as their father and for their father to report an occupation in both censuses.

Starting from this baseline sample, the minimum criteria for a father to be considered linked are a difference in age of no more than 3 years and a Jaro-Winkler distance of no more than 0.2 for first names (with an exception for initial-to-full name matches). We then drop “weak” links with a Jaro-Winkler distance for first names greater than 0.1, no initial match, an age difference greater than 1 year, and non-matching place of birth. A birth place is considered matched if either the county or parish of birth matches (allowing for one string to be contained in the other). Among remaining cases in census A (baseline) with more than one potential match in census B (target), we sequentially prioritize links with the following criteria:

1. Same county of birth
2. Same parish of birth
3. Same occupational string (allowing for one string to be contained in the other)
4. Closest match on age

We then repeat the above steps for records in census B with more than one potential link to census A. We keep individuals whose fathers are uniquely matched in both directions. At this

stage, the sample includes individuals whose father is uniquely linked from one census to the next. The remaining step considers cases in which there is more than one potential link for a set of siblings. Again, we repeat the above four steps to prioritize links and keep fathers who are uniquely matched in both directions. Our analysis excludes parishes with no hospital patients and drops households in which a patient is admitted to a hospital either before or after the linking window, but not within it.

## 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 conditions
2. Symptoms
3. Conditions requiring surgery
4. External factors (e.g. poisoning or collisions)
5. Foreign objects
6. Descriptors of severity
7. Body parts

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:

$$P(\text{Death in hospital})_{nha y}^g = \alpha + \theta_h + \delta_a + \gamma_y + \epsilon_{nha y}^g \quad (4)$$

where  $n^g$  indexes individual in-patient admissions for gender  $g$ ,  $h$  indexes hospitals,  $a$  indexes age in years at admissions, 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  $\widehat{\text{hospital}}(\theta)$ , age at admission ( $\delta$ ), and year of admission ( $\gamma$ ) fixed effects. We save the residuals  $\widehat{\epsilon}_{nha y}^g$  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.<sup>5</sup> 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:

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

which is the average unexplained mortality risk across all admissions 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.

## C Weighting

Here, we describe the procedure used to reweight the data by observable characteristics of all patients in the hospital records. The results are reported in Table A12 in the online appendix. Given that the hospital records represent the starting point for our sample construction, we take inpatients at risk of being linked to the census as the baseline population. The at risk population consists of patients born between 1870 and 1890 and admitted to one of the three hospitals in our sample at ages 0 to 11 between 1870 and 1902.

We re-weight the data to ensure that the final empirical samples match the proportions in the hospital records based on the following observable characteristics: age at admission, year at admission and place of residence. In specifications involving male and female patients we also re-weight by gender. Our procedure follows Abramitzky et al. (2020) and Black et al. (2020) in that we compute quintile bins for the continuous variables: age at admission (0-1, 2-3, 4-5, 6-8,

---

<sup>5</sup>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 thresholds of 10 or 50 observations.

9-11) and year of admission (1870-82, 1883-88, 189-93, 1894-98, 1899-02). Likewise, residential location is measured by the place of residence of a patient at the time of admission. We include indicators for registration districts of London inside and outside a hospital's catchment area, and counties of Greater London (Essex, Kent, Middlesex, and Surrey), with remaining counties as the excluded category. We estimate a probit regression on the population of hospital admissions with the dependent variable equal to one if an individual appears in the final empirical sample. Weights are computed as:

$$w = \frac{1}{\hat{p}} - 1$$

where  $\hat{p}$  is the predicted value from the probit regression. Intuitively, we assign higher weight to observations that are less likely to be matched.

## D Appendix Tables

Table A1: Number of beds and admissions to hospitals in London, 1894

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 six largest children's hospitals in London.

Source: [Chatto and Windus \(1897\)](#).

Table A2: Common causes of admission in hospital population and the final sample

Cause of admission	Hospital population			Cause of admission	Hospital male population			Cause of admission	Final sample	
	Frequency	Percent	Mortality rate		Frequency	Percent	Mortality rate		Frequency	Percent
Abscess	3,202	4.45	0.04	Abscess	1,901	4.52	0.04	Abscess	105	4.89
Diphtheria	2,773	3.85	0.38	Pneumonia	1,513	3.59	0.11	Pneumonia	65	3.03
Tubercular Disease	2,396	3.33	0.04	Diphtheria	1,499	3.56	0.36	Fracture	63	2.94
Pneumonia	2,368	3.29	0.11	Tubercular Disease	1,422	3.38	0.04	Bronchitis	57	2.66
Chorea	2,104	2.92	0.01	Bronchopneumonia	1,120	2.66	0.32	Phimosis	54	2.52
Bronchopneumonia	1,937	2.69	0.30	Bronchitis	999	2.37	0.17	Diphtheria	45	2.10
Bronchitis	1,766	2.45	0.16	Fracture	939	2.23	0.03	Chorea	44	2.05
Fracture	1,251	1.74	0.02	Meningitis	718	1.71	0.77	Typhoid Fever	42	1.96
Meningitis	1,226	1.70	0.77	Empyema	686	1.63	0.14	Tubercular Disease	40	1.86
Cleft Palate	1,093	1.52	0.00	Chorea	618	1.47	0.01	Injury	38	1.77
Empyema	1,090	1.51	0.13	Fever	613	1.46	0.11	Empyema	37	1.72
Typhoid Fever	1,006	1.40	0.06	Phimosis	601	1.43	0.01	Talipes	35	1.63
Fever	1,003	1.39	0.11	Injury	581	1.38	0.04	Cleft Palate	35	1.63
Tuberculosis	983	1.37	0.57	Typhoid Fever	569	1.35	0.07	Rheumatism	35	1.63
Morbus Cordis	954	1.33	0.17	Harelip	556	1.32	0.02	Rickets	32	1.49
Rheumatism	931	1.29	0.02	Tuberculosis	539	1.28	0.57	Necrosis	31	1.44
Harelip	852	1.18	0.02	Cleft Palate	535	1.27	0.01	Harelip	29	1.35
Talipes	799	1.11	0.01	Rheumatism	524	1.24	0.02	Morbus Cordis	28	1.30
Rickets	794	1.10	0.04	Talipes	517	1.23	0.00	Fever	28	1.30
Injury	776	1.08	0.03	Morbus Cordis	456	1.08	0.16	Scarlet Fever	25	1.16
Phthisis	775	1.08	0.23	Burn	451	1.07	0.25	Pleurisy	24	1.12
Burn	773	1.07	0.27	Rickets	440	1.05	0.05	Disease Knee	24	1.12
Diarrhea	713	0.99	0.26	Diarrhea	418	0.99	0.24	Nephritis	23	1.07
Disease Hip	704	0.98	0.02	Laryngitis	412	0.98	0.16	Disease Hip	21	0.98
Necrosis	679	0.94	0.03	Necrosis	409	0.97	0.03	Eczema	21	0.98
Total (top 25)	32,948	45.76	0.16	Total (top 25)	19,036	45.22	0.15	Total (top 25)	981	45.70
Outside top 25	39,006	54.24	0.11	Outside top 25	23,064	54.78	0.11	Outside top 25	1,165	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 2,146 hospital admissions reported in column 4 of Table 1, which correspond to the 1,849 male patients included in the main analysis in Table 2. The mortality rate is not shown for the final sample since it only includes patients who survived until adulthood.

Table A3: HISCLASS collapsed to four groups

White collar		Skilled		Semi-skilled		Unskilled	
1	Higher managers	6	Foremen	9	Lower skilled workers	10	Lower skilled farm workers
2	Higher professionals	7	Skilled workers			11	Unskilled workers
3	Lower managers	8	Farmers			12	Unskilled farm workers
4	Lower prof and clerical, sales						
5	Lower clerical and sales						

Notes: This table shows the Historical International Social Class Scheme (HISCLASS) ranks assigned to each of the four occupational rank categories used in the empirical analysis. HISCLASS is based on numeric code from the Historical International Standard Classification of Occupations (HISCO), which is a classification scheme developed for coding nineteenth century occupational titles ([Leeuwen et al. 2002](#)). The HISCLASS maps each of the 16,000 HISCO occupation codes to one of 12 social classes ranked from highest to lowest 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 ([van Leeuwen and Maas 2011](#)).

Table A4: 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	Police Constable	Bricklayer	House Painter	Farm Labourer
4	Commercial Clerk	French Polisher	Postman	Gardener Domestic
5	Shop Assistant	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 analysis of long-run occupational outcomes reported in Table 2. Column 1 combines professional, managerial and clerical occupations, which correspond to classes 1 to 5 in the Historical International Social Class Scheme (HISCLASS), into a white collar class. Column 2 subsumes farmers into skilled workers (HISCLASS 6 to 8), column 3 displays semi-skilled workers (HISCLASS 9), and column 4 combines unskilled workers as well as low and unskilled farm workers (HISCLASS 10 to 12). See section 3.3 for further details on the occupational classification.

Table A5: Intergenerational mobility matrix for linked population and estimation samples

	Father's occupational class				Total	N
	White collar	Skilled	Semi-skilled	Unskilled		
Panel A: Patients						
White collar	45.8	21.4	23.4	16.2	25.7	475
Skilled	21.1	38.7	24.1	22.8	27.7	512
Semi-skilled	22.2	26.9	38.2	31.9	30.4	563
Unskilled	11.0	13.0	14.3	29.1	16.2	299
N	356	561	568	364	1,849	
Panel B: Siblings						
White collar	49.7	23.4	28.9	16.8	28.8	533
Skilled	18.3	38.9	21.7	19.0	25.7	475
Semi-skilled	23.0	28.2	36.6	34.6	31.0	574
Unskilled	9.0	9.6	12.9	29.7	14.4	267
N	356	561	568	364	1,849	
Panel C: Patients and siblings						
White collar	47.8	22.4	26.1	16.5	27.3	1,008
Skilled	19.7	38.8	22.9	20.9	26.7	987
Semi-skilled	22.6	27.5	37.4	33.2	30.7	1,137
Unskilled	10.0	11.3	13.6	29.4	15.3	566
N	712	1,122	1,136	728	3,698	
Panel D: Population						
White collar	46.9	20.1	17.6	14.3	22.8	487,381
Skilled	20.7	40.6	19.9	19.5	26.0	555,840
Semi-skilled	22.8	26.1	50.0	31.5	34.4	737,389
Unskilled	9.5	13.2	12.4	34.7	16.8	360,607
N	376,200	617,898	694,080	453,039	2,141,217	

Notes: This table presents occupational transition matrices for fathers and sons in the main sample used in Table 2 (panels A to C) and in a synthetic population (panel D). Each column represents an occupational class for fathers of individuals in the main sample (from high to low): white collar, skilled, semi-skilled, and unskilled. The stratification of occupational titles into four groups is formed by consolidating the 12 strata of the Historical International Social Class Scheme (HISCLASS) as described in section 3.3. In panels A to C, the rows of the matrices represent the occupational class as adult for patients (panel A), siblings (panel B), or patients and siblings (panel C). The set of fathers does not change across the three panels. Within each panel, a cell contains the percentages of sons in each occupational class given the rank of the father. Percentages in a given column may not sum to 100 across its rows due to rounding errors. Panel D represents the transition matrix for a synthetic population constructed by sampling from five linked samples: 1881-1901, 1891-1901, 1881-1911, 1891-1911, and 1901-1911. The sampling probabilities correspond to the share of observations from each linked sample in the main sample. Each linked population consists of males born between 1870 and 1893 who were at least 18 years old in the later census year.

Table A6: Scaling of health estimates by intergenerational transmission of status

	Son's occupational status		
	(1) White collar	(2) Skilled +	(3) Ln wage
Panel A: Intergenerational occupational elasticities			
Father's status	0.250*** (0.001)	0.219*** (0.001)	0.246*** (0.001)
Mean of Y	0.240	0.488	4.586
N	2,053,932	2,053,932	2,055,293
Panel B: Scaled effects (%)			
Patient	17.3	16.5	13.0
Health deficiency index ( $\sigma$ )	7.3	6.8	5.0

Notes: Panel A presents estimates using data on males aged 0 to 11 linked from the 1881 to the 1911 complete-count census. In each column, the dependent variable is an indicator for the son's occupational status or the son's log occupational wage which are identical to the dependent variables in columns 3, 4, and 6 of Table 2. The treatment variable "father's status" varies across the columns and in each case is defined in an equivalent way as the dependent variable but for the father rather than the son. 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. Panel B displays percentages that represent the coefficients on the hospitalization indicator and a 1 s.d change in the health deficiency index from columns 3, 4, and 6 in Table 2 scaled by the estimates in panel A.

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



Table A7: Effects on long-run social outcomes

	(1)	(2)	(3)	(4)	(5)
	W. parents	Move county	Married	Any child	Ch. scholar
Panel A: Effects of hospital admission					
Patient	0.018 (0.014)	-0.017 (0.014)	-0.006 (0.014)	-0.018 (0.015)	0.031 (0.038)
% effect	5.0	5.7	1.2	4.8	9.7
Panel B: Effects of health deficiency index					
Health deficiency index	0.040 (0.045)	-0.047 (0.046)	0.015 (0.047)	-0.024 (0.049)	0.015 (0.123)
% effect ( $\sigma$ )	1.5	2.1	0.4	0.8	0.6
Mean of Y	0.354	0.293	0.508	0.374	0.318
N	3,698	3,698	3,684	3,698	660

Notes: This table presents sibling fixed effects estimates of the patient indicator (panel A) and the health deficiency index (panel B) on social outcomes observed at the time of census enumeration during adulthood. In column 1, the dependent variable is an indicator for living in the same household as at least one parent. In column 2, it is an indicator for residing in a different county than when enumerated in the census during childhood. In column 3, it is an indicator for being married at the time of enumeration, with or without a spouse present. In column 4, it is an indicator for having any children. In column 5, it is an indicator for having a child in school (see section 3.6 for details on how the variable is coded). In columns 1 to 5, the sample first restricts attention to individuals in the main empirical sample in Table 2, with further restriction to households in which the outcome variable is not missing for both patient and sibling. Column 5 also requires both patient and sibling to have a child. See Table 2 for a description of the control variables. Standard errors are clustered by childhood household. Percent effects in panel B are computed as  $\beta \times \sigma_{\text{HDI}}$  scaled by the dependent variable mean, where  $\sigma_{\text{HDI}} = 0.13$  is the standard deviation of the health deficiency index in the hospital population.

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

Table A8: Effects on disability in childhood and adulthood in expanded sample

	Childhood			Long Run
	(1) Males	(2) Females	(3) Both sexes	(4) Males
Panel A: Effects of hospital admission				
Patient	0.664*** (0.198)	1.070*** (0.240)	0.864*** (0.153)	0.676*** (0.213)
% effect	153.1	179.2	169.7	117.4
Panel B: Effects of health deficiency index				
Health deficiency index	2.085*** (0.665)	2.778*** (0.698)	2.498*** (0.503)	1.710** (0.707)
% effect ( $\sigma$ )	62.1	60.0	63.3	38.3
Mean of Y	0.434	0.597	0.509	0.576
N	8,534	7,370	15,904	5,038

Notes: This table presents sibling fixed effects estimates from specifications in which the dependent variable is an indicator variable equal to one if an individual is recorded as having a disability in the census. We multiply the dependent variable by 100 for easier interpretation of magnitudes. Columns 1 to 3 present estimates for childhood disability, which is recorded in the census up to 10 years post hospital admission. Column 4 shows estimates for disability in adulthood and involves linkages to two censuses. In comparison to the main result for disability reported in Table 3, in which the sample is restricted to patients born between 1870 and 1890, the sample here includes patients born up to 1902. See section 3.7 for a description of how the variable is coded and for a discussion of the sample construction. In panel A, the explanatory variable of interest is an indicator for hospital admission, and in panel B, it is the health deficiency index. See Table 2 for a description of the control variables. Standard errors are clustered by childhood household. Percent effects in panel B are computed as  $\beta \times \sigma_{\text{HDI}}$  scaled by the dependent variable mean, where  $\sigma_{\text{HDI}} = 0.13$  is the standard deviation of the health deficiency index in the hospital population.

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

Table A9: Heterogeneity by severity and age at admission

	(1) Class ↗	(2) Class ↘	(3) White collar	(4) Skilled +	(5) Unskilled	(6) Log wage
Panel A: Interaction with above vs. below median HDI						
Patient × low-HDI	-0.031* (0.017)	0.027 (0.016)	-0.042** (0.018)	-0.030 (0.020)	0.028* (0.015)	-0.038** (0.015)
Patient × high-HDI	-0.030 (0.020)	0.029 (0.018)	-0.045** (0.020)	-0.045* (0.024)	0.039** (0.017)	-0.025 (0.017)
P-value	0.986	0.921	0.917	0.608	0.608	0.549
Panel B: Interaction with early (0-4) vs. late (5-11) childhood admission						
Patient × [0-4]	-0.059*** (0.021)	0.022 (0.020)	-0.056** (0.023)	-0.057** (0.026)	0.041** (0.018)	-0.045** (0.019)
Patient × [5-11]	-0.020 (0.017)	0.029* (0.016)	-0.041** (0.017)	-0.025 (0.020)	0.029** (0.014)	-0.026* (0.014)
P-value	0.134	0.801	0.584	0.303	0.579	0.397
Mean of Y	0.357	0.260	0.273	0.539	0.153	4.628
N	3,698	3,698	3,698	3,698	3,698	3,698

Notes: This table presents sibling fixed effects estimates. Panel A displays estimates in which we interact the indicator for hospital patient with indicators for being admitted for conditions with above and below median values of the 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. The dependent variables in columns 1 to 6 correspond to those shown in Table 2. See Table 2 for a description of the 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 A10: Descriptive statistics for hospital catchment areas

	(1)	(2)	(3)	(4)
	Barts	GOSH	Guys	Rest of London
Share age 0 to 4	0.121	0.110	0.131	0.119
Share age 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. The 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 A11: Selection into hospitalization for male patients

	Observed in hospital records [ $\times 100$ ]		
	(1)	(2)	(3)
Father skilled	0.061*** (0.020)	0.039 (0.031)	0.038 (0.024)
Father semi-skilled	0.069*** (0.020)	0.053* (0.027)	0.053** (0.022)
Father unskilled	0.089*** (0.025)	0.074* (0.037)	0.071** (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: This table presents OLS estimates using a sample that 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.

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

Table A12: Weighting by characteristics of patient population

	Occupational						Schooling	Disability		
	(1) Class ↗	(2) Class ↘	(3) White collar	(4) Skilled +	(5) Unskilled	(6) Log wage	(7) Scholar	(8) Pre-existing	(9) Childhood	(10) Long-run
Panel A: Effects of hospital admission										
Patient	-0.030** (0.014)	0.025* (0.013)	-0.043*** (0.015)	-0.036** (0.017)	0.027** (0.012)	-0.034*** (0.012)	-0.023*** (0.008)	0.090 (0.116)	0.772*** (0.190)	0.810*** (0.244)
% effect	8.5	9.7	15.7	6.8	17.6	3.4	3.1	82.3	149.3	136.7
Panel B: Effects of health deficiency index										
Health deficiency index	-0.104** (0.047)	0.073* (0.045)	-0.141*** (0.048)	-0.115** (0.056)	0.089** (0.039)	-0.104*** (0.039)	-0.050 (0.034)	0.065 (0.268)	2.385*** (0.604)	2.013** (0.848)
% effect ( $\sigma$ )	3.8	3.6	6.7	2.8	7.4	1.3	0.9	7.6	59.5	43.9
Mean of Y	0.355	0.260	0.273	0.538	0.155	4.629	0.729	0.109	0.517	0.592
N	3,698	3,698	3,698	3,698	3,698	3,698	3,040	9,096	11,312	4,312

Notes: This table displays sibling fixed effects estimates from regressions that re-weight the data by the following observable characteristics in the hospital records: admission age, admission year, and location of residence in all columns, as well as gender in columns 7 to 9. Re-weighting the data ensures that each linked sample matches the proportions in the population of inpatient hospital admissions. The weighting procedure is described in more detail in online appendix C. In panel A, the explanatory variable of interest is an indicator for hospital admission, and in panel B, it is the health deficiency index. Columns 1 to 6 present estimates for the outcomes variables shown in Table 2. Columns 7 to 9 display estimates for the schooling and disability mechanisms from the pooled gender samples in Table 3, while column 10 reports estimates for long-run disability using the linked sample of boys only. See Table 2 for a description of the control variables. Standard errors are clustered by childhood household. Percent effects for log wages (column 6) are calculated using the formula  $100 \times \exp(\beta) - 1$ . Percent effects in panel B are computed as  $\beta \times \sigma_{\text{HDI}}$  scaled by the dependent variable mean, where  $\sigma_{\text{HDI}} = 0.13$  is the standard deviation of the health deficiency index in the hospital population.

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)	(2)	(3)	(4)	(5)	(6)
	Baseline estimate	Drop high mortality	Drop low mortality	Drop infant admission	Drop multiple admissions	Drop contagious
Panel A: Effects on P(Class ↗)						
Health deficiency index	-0.098** (0.046)	-0.135*** (0.051)	-0.097** (0.047)	-0.091* (0.048)	-0.116** (0.049)	-0.141*** (0.053)
Mean of Y	0.357	0.356	0.355	0.359	0.357	0.356
Panel B: Effects on P(Class ↘)						
Health deficiency index	0.084* (0.043)	0.083* (0.049)	0.091** (0.044)	0.072 (0.045)	0.086* (0.046)	0.079 (0.051)
Mean of Y	0.260	0.256	0.261	0.261	0.260	0.263
Panel C: Effects on P(White collar)						
Health deficiency index	-0.142*** (0.046)	-0.164*** (0.053)	-0.146*** (0.047)	-0.158*** (0.048)	-0.171*** (0.049)	-0.181*** (0.054)
Mean of Y	0.273	0.269	0.278	0.276	0.267	0.276
Panel D: Effects on P(Skilled +)						
Health deficiency index	-0.116** (0.054)	-0.149** (0.061)	-0.118** (0.055)	-0.099* (0.056)	-0.126** (0.057)	-0.134** (0.063)
Mean of Y	0.539	0.538	0.541	0.546	0.534	0.541
Panel E: Effects on P(Unskilled)						
Health deficiency index	0.103*** (0.039)	0.110** (0.043)	0.095** (0.039)	0.102** (0.040)	0.108*** (0.042)	0.126*** (0.046)
Mean of Y	0.153	0.155	0.151	0.153	0.152	0.154
Panel F: Effects on log occupational wage						
Health deficiency index	-0.096** (0.038)	-0.120*** (0.045)	-0.094** (0.039)	-0.116*** (0.040)	-0.118*** (0.041)	-0.139*** (0.046)
Mean of Y	4.628	4.627	4.629	4.630	4.624	4.634
N	3,698	3,328	3,328	3,350	3,256	3,044

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(Class ↗)							
Health deficiency index	-0.098** (0.046)	-0.069 (0.042)	-0.066 (0.042)	-0.107** (0.050)	-0.095** (0.048)	-0.111** (0.048)	-0.119** (0.055)
Mean of Y	0.357	0.355	0.355	0.367	0.357	0.353	0.363
Panel B: Effects on P(Class ↘)							
Health deficiency index	0.084* (0.043)	0.078* (0.040)	0.072* (0.040)	0.084* (0.046)	0.097** (0.046)	0.064 (0.045)	0.073 (0.053)
Mean of Y	0.260	0.259	0.260	0.265	0.261	0.258	0.267
Panel C: Effects on P(White collar)							
Health deficiency index	-0.142*** (0.046)	-0.120*** (0.043)	-0.117*** (0.042)	-0.141*** (0.050)	-0.152*** (0.049)	-0.153*** (0.049)	-0.167*** (0.057)
Mean of Y	0.273	0.275	0.274	0.278	0.279	0.276	0.286
N	3,698	4,473	4,512	3,032	3,332	3,302	2,366
Panel D: Effects on P(Skilled +)							
Health deficiency index	-0.116** (0.054)	-0.089* (0.050)	-0.081 (0.050)	-0.125** (0.059)	-0.145** (0.058)	-0.122** (0.057)	-0.132** (0.066)
Mean of Y	0.539	0.539	0.539	0.546	0.550	0.540	0.550
Panel E: Effects on P(Unskilled)							
Health deficiency index	0.103*** (0.039)	0.070* (0.036)	0.071** (0.036)	0.119*** (0.042)	0.099** (0.041)	0.084** (0.041)	0.073 (0.048)
Mean of Y	0.153	0.156	0.155	0.144	0.145	0.154	0.142
Panel F: Effects of log occupational wage							
Health deficiency index	-0.096** (0.038)	-0.075** (0.036)	-0.075** (0.035)	-0.085** (0.041)	-0.104** (0.041)	-0.103** (0.041)	-0.120** (0.048)
Mean of Y	4.628	4.629	4.629	4.639	4.634	4.628	4.650
N	3,698	4,473	4,512	3,032	3,332	3,302	2,366

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: Long-run outcomes: Robustness to variations in occupational status

	(1) Baseline estimate	(2) Impute Hhld SES	(3) Highest Hhld SES	(4) High class if missing	(5) Low class if missing
Panel A: Effects on P(Class ↗)					
Health deficiency index	-0.098** (0.046)	-0.102** (0.044)	-0.089** (0.045)	-0.072 (0.044)	-0.058 (0.046)
Mean of Y	0.357	0.360	0.346	0.339	0.386
Panel B: Effects on P(Class ↘)					
Health deficiency index	0.084* (0.043)	0.085** (0.042)	0.082* (0.043)	0.081* (0.044)	0.067 (0.042)
Mean of Y	0.260	0.258	0.273	0.297	0.251
Panel C: Effects on P(White collar)					
Health deficiency index	-0.142*** (0.046)	-0.133*** (0.045)	-0.133*** (0.045)	-0.130*** (0.044)	-0.116** (0.047)
Mean of Y	0.273	0.270	0.270	0.259	0.306
Panel D: Effects on P(Skilled +)					
Health deficiency index	-0.116** (0.054)	-0.120** (0.053)	-0.120** (0.053)	-0.105** (0.052)	-0.091* (0.053)
Mean of Y	0.539	0.542	0.542	0.511	0.558
Panel E: Effects on P(Unskilled)					
Health deficiency index	0.103*** (0.039)	0.103*** (0.038)	0.103*** (0.038)	0.091** (0.037)	0.105** (0.041)
Mean of Y	0.153	0.151	0.151	0.147	0.194
N	3,698	3,870	3,870	4,004	4,004

Notes: This table is identical to Table 6 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 A16: Robustness in schooling outcomes

	(1) Column (3) 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	-0.026*** (0.008)	-0.023*** (0.009)	-0.025*** (0.009)	-0.025*** (0.009)	-0.029*** (0.009)	-0.018** (0.009)
Health deficiency index	-0.079*** (0.028)	-0.070** (0.032)	-0.081*** (0.029)	-0.079*** (0.030)	-0.091*** (0.031)	-0.047 (0.031)
Mean of Y	0.743	0.755	0.745	0.750	0.751	0.748
N	3,040	2,728	2,698	2,684	2,672	2,504
	(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	-0.028*** (0.008)	-0.027*** (0.008)	-0.024** (0.010)	-0.030*** (0.009)	-0.019** (0.008)	-0.017* (0.010)
Health deficiency index	-0.083*** (0.026)	-0.080*** (0.026)	-0.070** (0.030)	-0.092*** (0.031)	-0.068** (0.029)	-0.061* (0.033)
Mean of Y	0.744	0.745	0.732	0.744	0.748	0.726
N	3,340	3,363	2,418	2,774	2,784	1,980

Notes: Each cell comes from a separate sibling fixed effects regression. In panel A, the explanatory variable of interest is an indicator for hospital admission, and in panel B, it is the health deficiency index. Column 1 of panel A reproduces the estimate for the schooling outcome from column 3 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 A17: Robustness of pre-existing disability outcome

	(1) Column (6) 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	0.101 (0.118)	0.104 (0.129)	0.102 (0.132)	0.049 (0.113)	0.067 (0.099)	0.198 (0.140)
Health deficiency index	0.029 (0.303)	-0.023 (0.432)	0.123 (0.320)	-0.082 (0.304)	-0.060 (0.269)	0.235 (0.435)
Mean of Y	0.154	0.171	0.160	0.151	0.140	0.171
N	9,096	8,194	8,128	8,612	7,874	7,018
	(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	0.101 (0.116)	0.100 (0.113)	0.072 (0.127)	0.134 (0.132)	0.048 (0.124)	0.044 (0.158)
Health deficiency index	0.054 (0.280)	0.059 (0.271)	-0.082 (0.328)	0.095 (0.360)	-0.130 (0.324)	-0.239 (0.422)
Mean of Y	0.153	0.150	0.151	0.161	0.147	0.204
N	12,420	12,637	7,268	8,058	8,174	5,892

Notes: Each cell comes from a separate sibling fixed effects regression. In panel A, the explanatory variable of interest is an indicator for hospital admission, and in panel B, it is the health deficiency index. Column 1 of panel A reproduces the estimate for the pre-existing disability outcome from column 6 of Table 3. 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 A18: Robustness of childhood disability outcome

	(1) Column (9) 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	0.746*** (0.185)	0.750*** (0.200)	0.719*** (0.197)	0.828*** (0.208)	0.705*** (0.193)	0.726*** (0.210)
Health deficiency index	2.235*** (0.569)	2.392*** (0.670)	2.148*** (0.599)	2.325*** (0.615)	1.955*** (0.559)	2.141*** (0.670)
Mean of Y	0.513	0.531	0.542	0.545	0.474	0.535
N	11,312	10,178	10,154	10,098	9,912	9,160
	(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	0.716*** (0.165)	0.713*** (0.161)	0.573*** (0.196)	0.794*** (0.195)	0.761*** (0.194)	0.974*** (0.254)
Health deficiency index	2.170*** (0.543)	2.147*** (0.528)	1.772*** (0.607)	2.277*** (0.597)	2.287*** (0.600)	2.899*** (0.773)
Mean of Y	0.422	0.420	0.454	0.529	0.526	0.671
N	16,346	16,656	9,030	10,202	10,270	7,454

Notes: Each cell comes from a separate sibling fixed effects regression. In panel A, the explanatory variable of interest is an indicator for hospital admission, and in panel B, it is the health deficiency index. Column 1 of panel A reproduces the estimate for the childhood disability outcome from column 9 of Table 3. 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 A19: Robustness in long-run disability outcomes

	(1) Column (10) 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	0.739*** (0.233)	0.853*** (0.248)	0.764*** (0.252)	0.730*** (0.239)	0.701*** (0.253)	0.920*** (0.250)
Health deficiency index	1.866** (0.799)	2.815*** (0.897)	1.949** (0.805)	1.674** (0.791)	1.783** (0.897)	3.106*** (0.908)
Mean of Y	0.580	0.593	0.593	0.562	0.554	0.567
N	4,312	3,880	3,880	3,918	3,790	3,530
	(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	0.560** (0.240)	0.559** (0.237)	0.651** (0.260)	0.841*** (0.254)	0.815*** (0.238)	1.042*** (0.308)
Health deficiency index	1.416* (0.818)	1.408* (0.806)	1.412 (0.863)	2.049** (0.883)	2.164*** (0.714)	2.741*** (0.949)
Mean of Y	0.484	0.479	0.569	0.593	0.544	0.576
N	4,547	4,591	3,512	3,878	3,862	2,776

Notes: Each cell comes from a separate sibling fixed effects regression. In panel A, the explanatory variable of interest is an indicator for hospital admission, and in panel B, it is the health deficiency index. Column 1 of panel A reproduces the estimate for the long-run disability outcome from column 10 of Table 3. 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 A20: Sibling-specific determinants of hospitalization

	Hospitalization (Patients vs. siblings)				Health deficiency index   Hospitalization (Patients only)			
	(1) 1881	(2) 1891	(3) 1901	(4) Any	(5) 1881	(6) 1891	(7) 1901	(8) Any
First born	-0.073* (0.039)	-0.158*** (0.039)	-0.054 (0.101)	-0.109*** (0.027)	-0.004 (0.008)	-0.002 (0.007)	-0.001 (0.015)	-0.003 (0.005)
Female	-0.043** (0.020)	-0.063*** (0.020)	-0.101** (0.045)	-0.057*** (0.013)	-0.000 (0.006)	-0.001 (0.006)	0.001 (0.014)	-0.001 (0.004)
First born × female	0.123** (0.056)	0.122** (0.053)	0.007 (0.133)	0.110*** (0.037)	-0.006 (0.012)	0.012 (0.011)	-0.002 (0.026)	0.002 (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 consist 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 A21: Patient health deficiency index and likelihood of linkage to census

	Census year linked to hospital records			
	(1) 1881	(2) 1891	(3) 1901	(4) Any
Health deficiency index	-0.097 (0.069)	-0.019 (0.038)	-0.090** (0.040)	-0.059** (0.026)
Mean of Y	0.265	0.272	0.251	0.263
N	4,758	12,355	9,772	26,885

Notes: This table presents OLS estimates from specifications in which the dependent variable is an indicator for a unique match from the hospital records to a census. The only explanatory variable is the health deficiency index, the coefficients for which are reported in the table. Columns 1 to 3 present results for linkages from the hospital records to the 1881, 1891, and 1901 censuses, respectively. Column 4 pools together the samples in the preceding columns. 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.

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

Table A22: Census-to-census linkage rates

	Outcome year = 1901		Outcome year = 1911		
	(1) 1881	(2) 1891	(3) 1881	(4) 1891	(5) 1901
No match	0.336	0.267	0.368	0.319	0.243
Multiple matches	0.127	0.136	0.122	0.119	0.114
Unique match	0.537	0.598	0.510	0.562	0.643
Baseline sample	5,864,701	6,425,991	5,864,701	6,425,991	6,764,357

Notes: The table presents census-to-census linkage rates for boys residing in England in the base-year census. See appendix A.1 for a description of the linkage procedure.

Table A23: Linkage rates to census records by hospital (males)

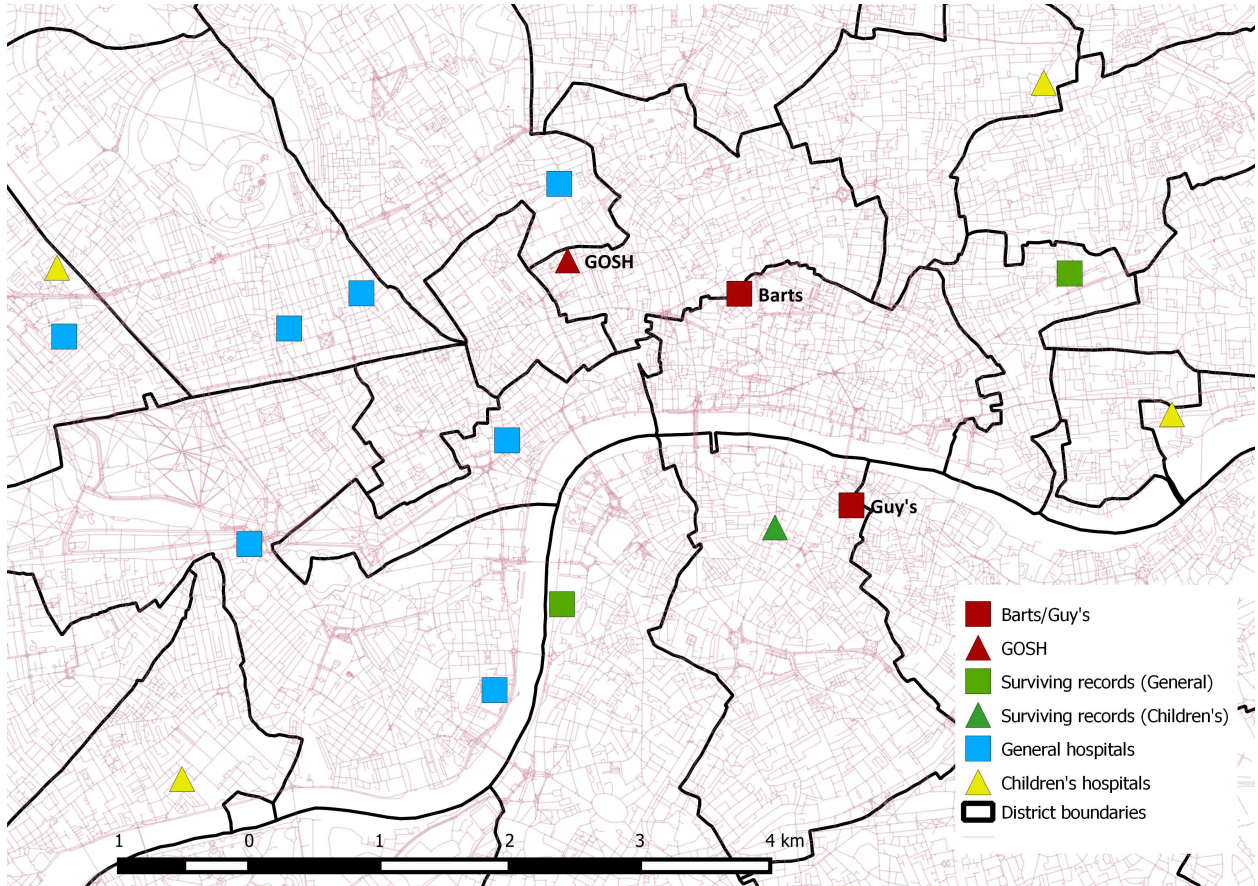
	Census year linked to hospital records			
	(1) 1881	(2) 1891	(3) 1901	(4) Any
Panel A: Barts hospital				
No match	0.063	0.050	0.155	0.031
Multiple matches	0.664	0.660	0.582	0.616
Unique match	0.273	0.290	0.263	0.353
Sibling present	0.257	0.264	0.222	0.334
Patient matched	0.192	0.172	0.143	0.232
Patient and sibling	0.077	0.074	0.036	0.117
In final sample	0.058	0.046	0.023	0.080
Total admissions	8,441	11,551	4,850	14,221
Panel B: GOSH for Sick Children				
No match	0.069	0.064	0.212	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	0.055	0.042	0.022	0.071
Total admissions	6,052	9,040	4,522	11,317
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	0.055	0.037	0.022	0.069
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 for males, separately by hospital. See Table 1 for a description of each sample restriction.



## E Appendix Figures

Figure A1: London hospital locations and surviving records



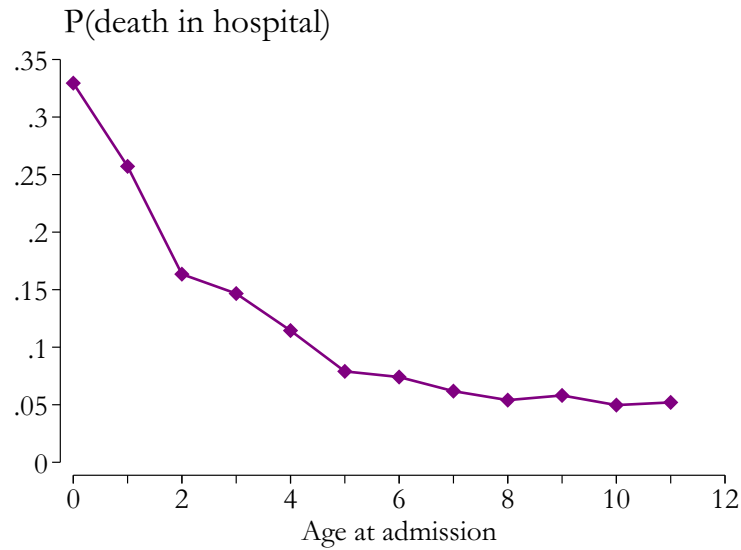
Notes: A map of central London marking the locations of hospitals in the empirical sample (red squares and triangle), general hospitals (squares symbol), and children's hospitals (triangle symbol). The subset of these hospitals with surviving archival records is marked in green.

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

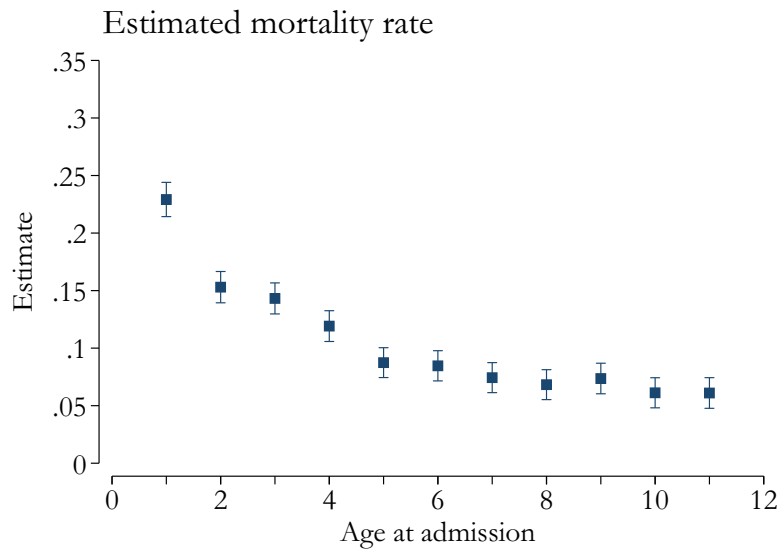
No.	DATE OF ADMISSION, 1870.	NAME.	AGE.	COMPLAINT.	ADDRESS.	WARD.	PHYSICIAN OR SURGEON.	DATE OF DISCHARGE OR DEATH.
3526	Nov 7	Henry Walker	57	Pott's pnc	31 Moreland Street, City Road	Colston	W. Longton	14 Dec.
3527		Stephen Kent	47	etc	76 Lake Road, Landport, Portsmouth	Commonly	W. Blackwood	10 Dec.
3528		Charles Chesterton	15	Abn. of hand	2 <sup>nd</sup> Dunnington Road, High St. Hampton	Abaker	W. Smith	21 Dec.
3529		Robert Kitting	62	Swollen leg	33 Grove Road, Gray's Court			22 Nov.
3530		James Stephens	6	Call on head	6 Underwood Road, Shepherdia Walk	Abaker	W. Hillitt	26 Dec.
3531		Thomas Jarwood	16	Chinura	18 White Street, Commercial Road	Logans	W. Longton	6 Dec.
3532	11	John Parker Hayling	24	St	16 J. Block, Rose Street, Leek			27 Nov.
3533		Henry Anderson	25	Hallux valgus	Stone cot., Sutton Common, Surrey	Abaker	W. Hillitt	6 Dec.
3534		William Charles Jones	18	Pain in chest	11 St. Helena Place, Camberwell	Matthews	D. Church	12 Dec.
3535		Ernest Francis Taylor	14	Small tumour on back	23 Frankham, E.C.	Henry	W. Smith	12 Dec.
3536		John Paterson	38	Gon & Expt.	8 Gunpowder Alley, E.C.	Logans	W. Longton	13 Dec.
3537		James John White	19	Delirium	70 Cadogan Place, S.W.	Leek	D. Gee	7 Dec.
3538		Arthur Smith	7	Tuber. fungus on thumb	Headlight Villa, Newnham Road, Wood Green	Henry	W. Longton	21 Dec.
3539		John Mitchell	43	Cerebral Haem.	15 Pleasant Place, Essex Road	John	W. Blackwood	12 Dec.
3540		John Spence	22	Acute Strabismus of eye	91 East Street, Watford	Darker	W. Longton	16 Dec.
3541		Frederick Tucker	42	Haematococis	4 Princess Villas, Princess St. Condon	Leek	D. Gee	18 Nov.
3542		Edwin Jones	65	Conjunctivitis of eye		Commonly	W. Longton	15 Nov.
3543		Joseph Gartham	30	Pott's pnc	34 White Lion Street, Pentonville	Colston		27 Nov.
3544		Charles Fisher	16	Scalp Wound	31 Turner Street, Cranford Heath			1 Jan.
3545		Henry Dennis	40	Tract. Pellicla	33 Newbark Street, New North Road			1 Jan.
3546	12	William Augustus Barnard	19	Division of the heart	4 Crooked Row, Chancery Lane, Strand	Abaker	W. Hillitt	2 Jan.
3547		Eliza Webb	47	Mor. Corpore	73 Abbot Hall, Blackfriars	Leek	D. Gee	25 Dec.
3548		John Clark	27	Abcess of eye	8 Cannon Terrace, Mile End Road	Matthews	D. Church	15 Jan.
3549		John W. Gane	32	Traumatic Cataract	77 Long Street, Kingsland Road	Optic	W. Paves	9 Jan.
3550		Charles Thompson & Co.	45	Phasia	2 Elgin Villas, Myddleton Road, South Park	Leek	D. Gee	3 Dec.
3551	10/11/70	Henry Rodgers	40	Securis of upper jaw	7 Myrtle Alley, New Street, Melrose	Leek	W. Longton	6 Dec.
3552		Henry Betty	40	Securis of upper jaw	4 Davies Rd. 10 <sup>th</sup> Avenue, North Commercial Rd.	Abaker	W. Hillitt	5 Jan.
3553		William Bennett	23	Strum. dis. of throat	The Liverpool, Birmingham Hotel, Custom	Henry	W. Smith	25 Nov.
3554		David Ayloth	51	Gon. Thromb.	42 Forest Road, Dalston	Logans	W. Longton	2 Nov.
3555		James Christie	32	Haematococis	328 Oxford Street, W.	John	W. Blackwood	26 Dec.
3556		Henry Mason	46	Paraplegia	21 East Street, Hatford	Mark	D. Andrew	14 Jan.
3557		George Kennard	37	Malocclus	5 Bridborough Street, Fiddlers Hall			10 Dec.
3558		Ernest Vyryan	16	Strum. of neck	59 Durham Road, Stockwell Road	Abaker	W. Hillitt	26 Dec.
3559		William Leonard	12	Spina. of neck	138 Conhal Street, Leek			6 Dec.
3560		Frank Calver Abbott	11	Strum. of thigh	Portland Villa, Grange Park Road, Leyton			31 Dec.
3561		William Long	49	Spina. of neck	21 Bath Street, Hereford			2 Jan.
3562		John Barber	24	Tract. base skull	11 Langton Avenue, Langton Street, Leek, Commonly	W. Longton		16 Dec.
3563	13	William Allen Knight	30	Thromb. of	185 Bye Lane, Beckham	Logans		1 Dec.
3564		Arthur James Fuller	35	Chronic effusion	8 Salisbury Court, Bethnal Green Road	Mark	D. Andrew	12 Dec.
3565		Matthew Mudd	24	Abcess of hand	67 Woodstock Street, Fanning Town			20 Nov.
3566		Alexander Knight	76	Paraplegia	91 Grosvenor Park, Camberwell			17 Dec.

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



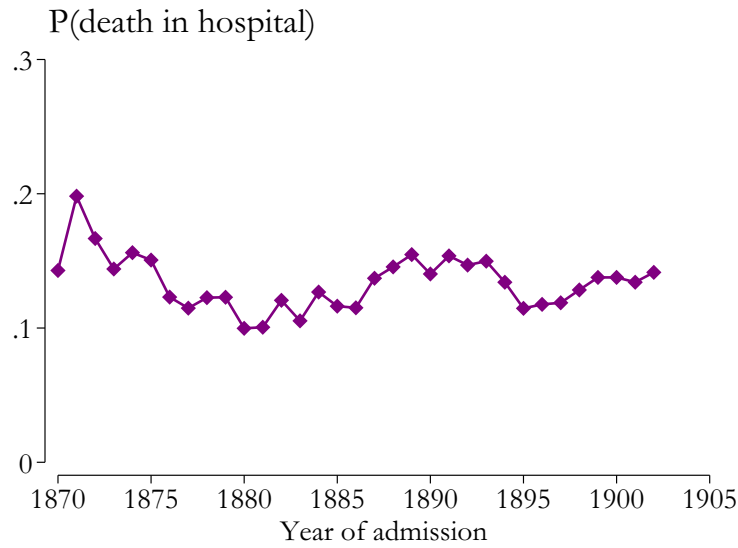
(a) Raw data



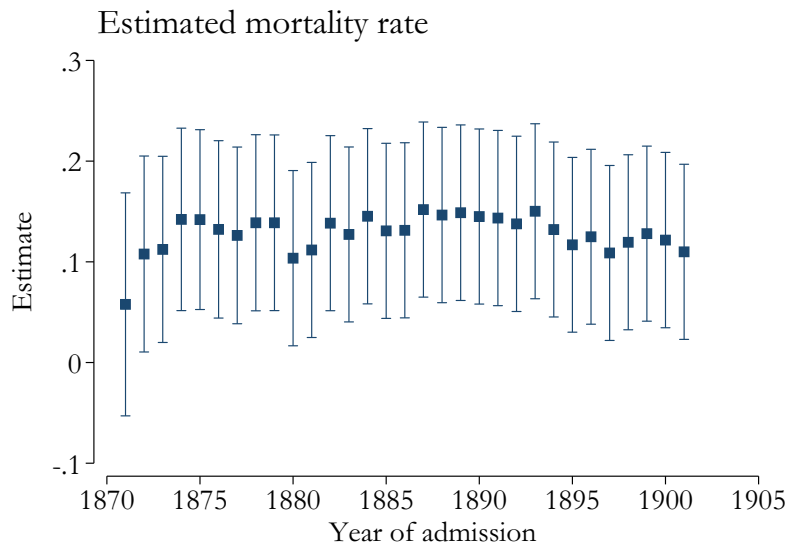
(b) Regression adjusted

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. The samples include data on all in-patients aged 0 to 11 born between 1870 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



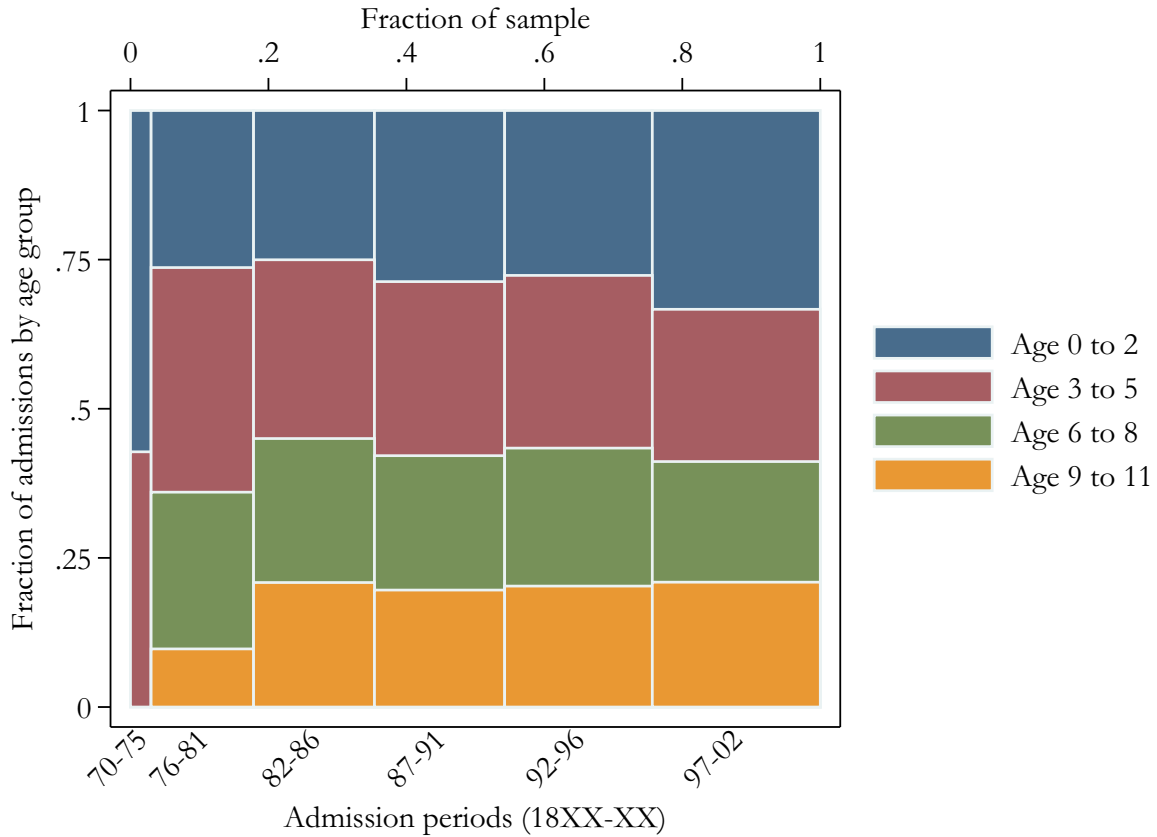
(a) Raw data



(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 1870 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.

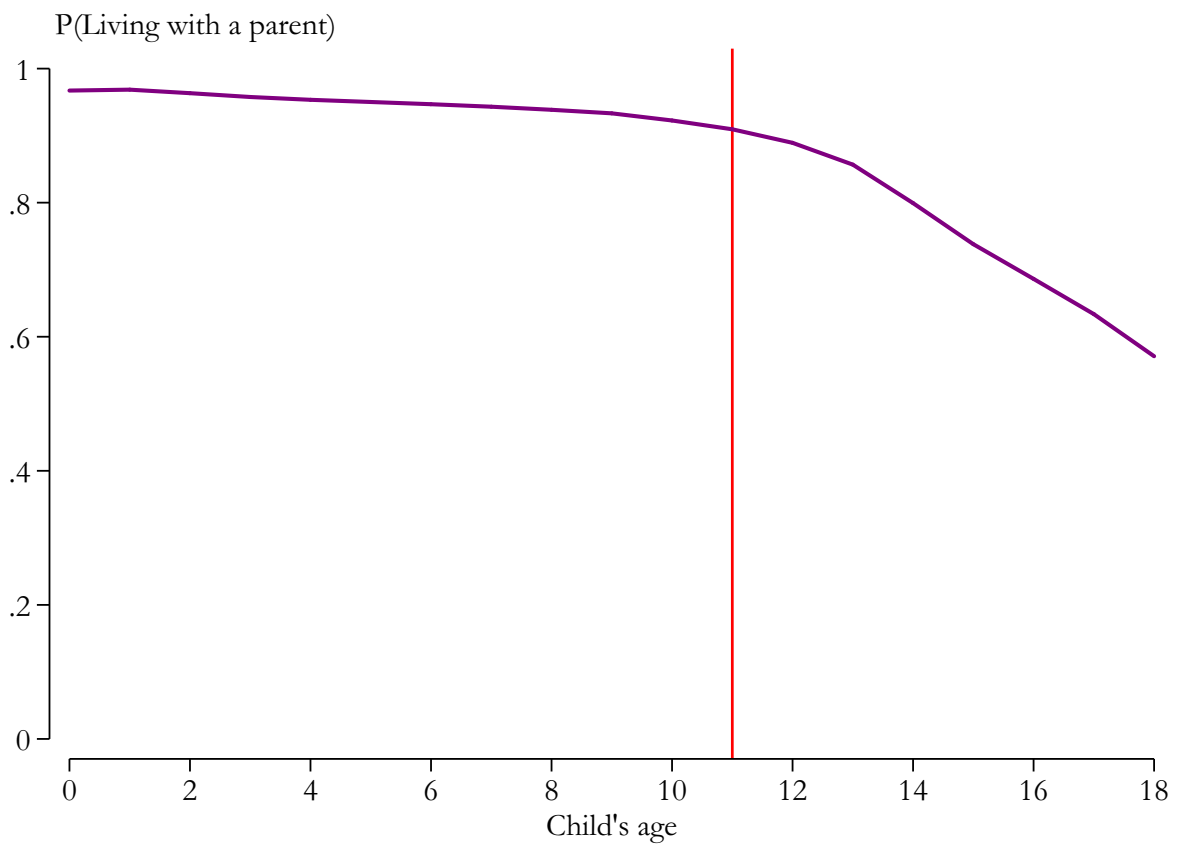
Figure A5: Admissions by age group and period



Notes: This figure provides a visualization of in-patient admissions by age groups (0 to 2, 3 to 5, 6 to 8, and 9 to 11) and admission periods (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.

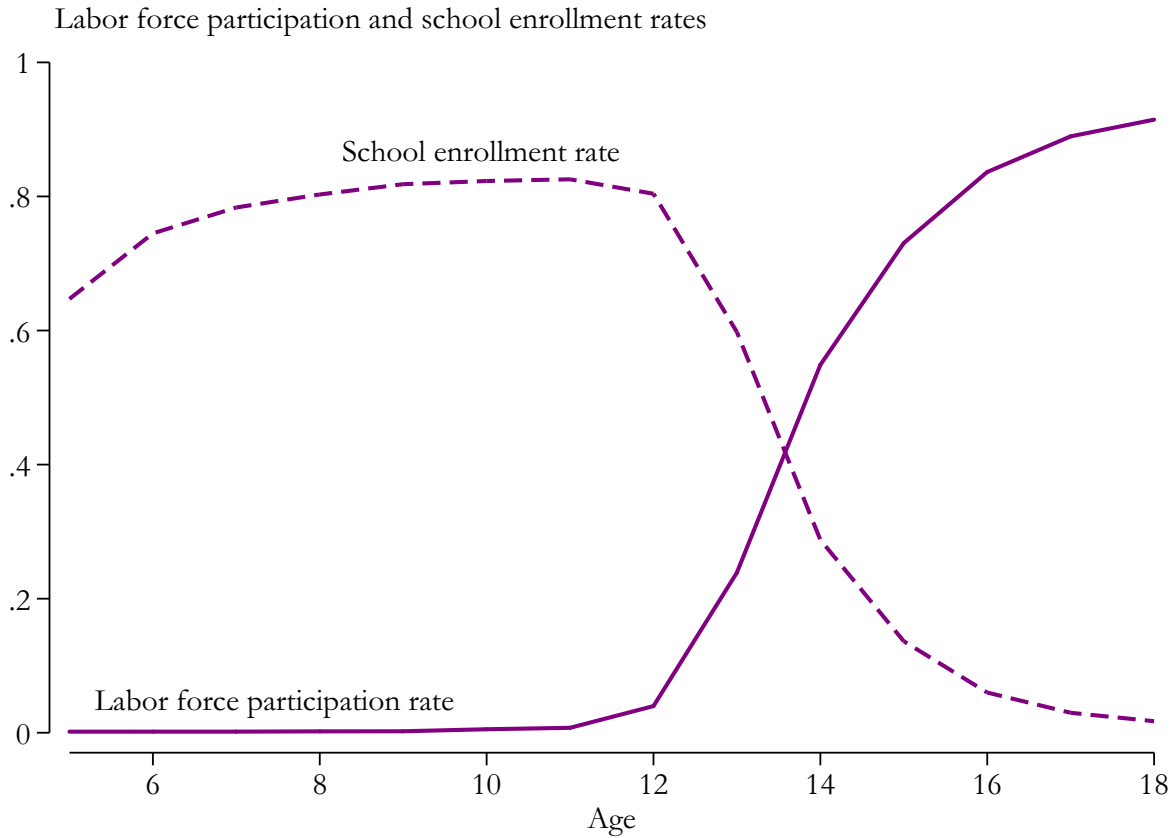


Figure A6: Share of children living with a parent in 1881, by age



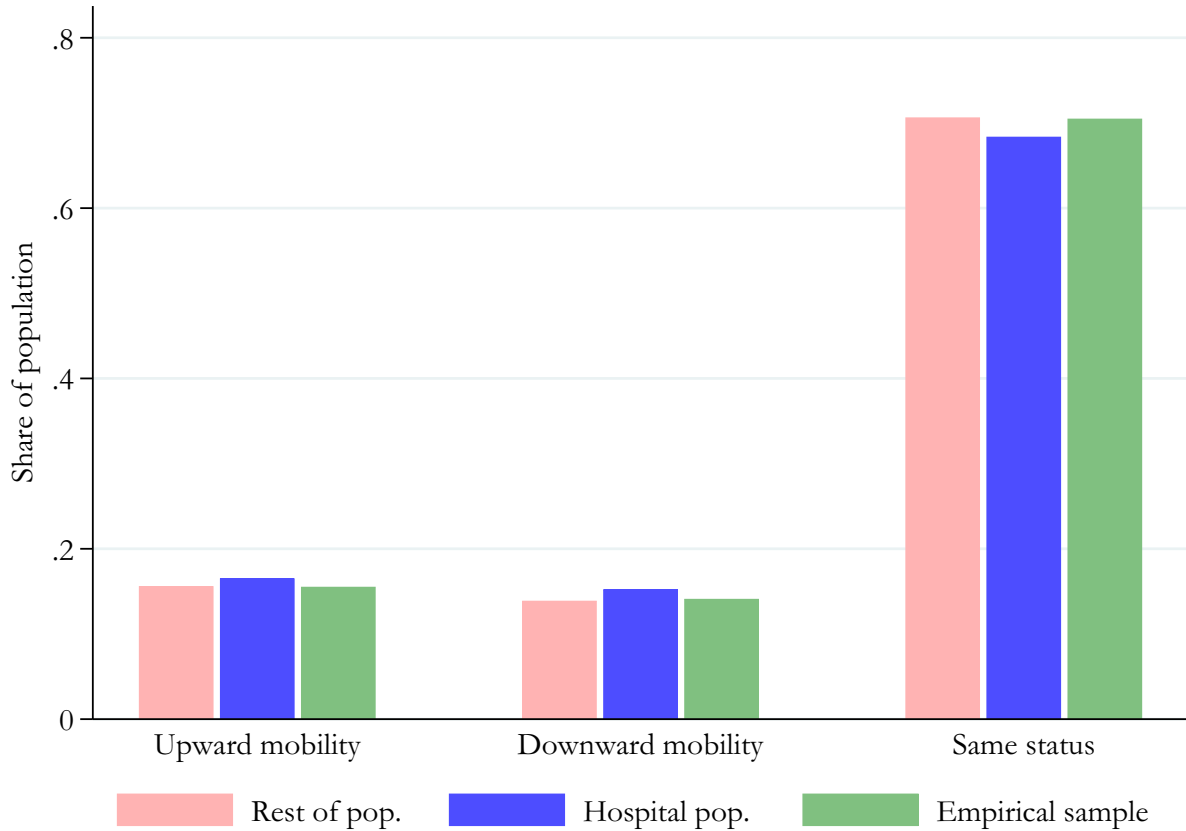
Notes: This figure plots the share of children who were enumerated in the 1881 census in a household in which at least one parent was present. The sample consists of all households in the County of London. Results are similar in the 1891 census.

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 to be in the labor force if any gainful occupation is reported in the census as measured by a valid Historical International Standard Classification of Occupations (HISCO) code. See section 3.6 for a description of how school enrollment is coded. Results are similar in the 1891 census.

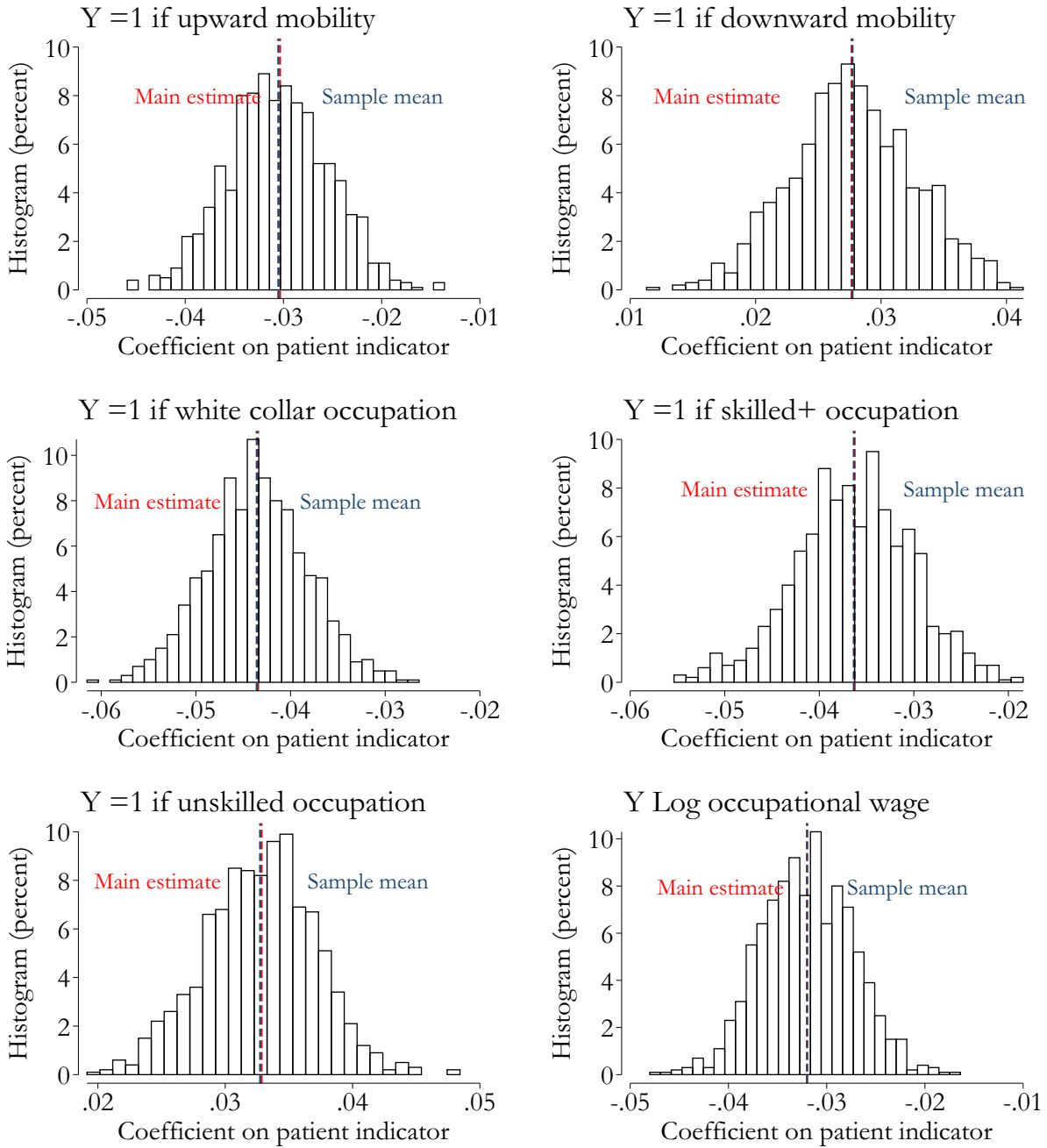
Figure A8: Changes in father's occupation across adjacent censuses



Notes: This figure presents summary measures of occupational changes for fathers linked between the 1881 and 1891 or 1891 and 1901 censuses. It shows the share of fathers whose occupational rank increases (upward mobility), decreases (downward mobility) or stays the same (same status) from one census to the next. Shares are shown separately for the subset of the main empirical sample with linked fathers (green), the rest of the hospital population (blue), and the rest of the overall population (red). See appendix A.5 for a description of the procedure for linking fathers.

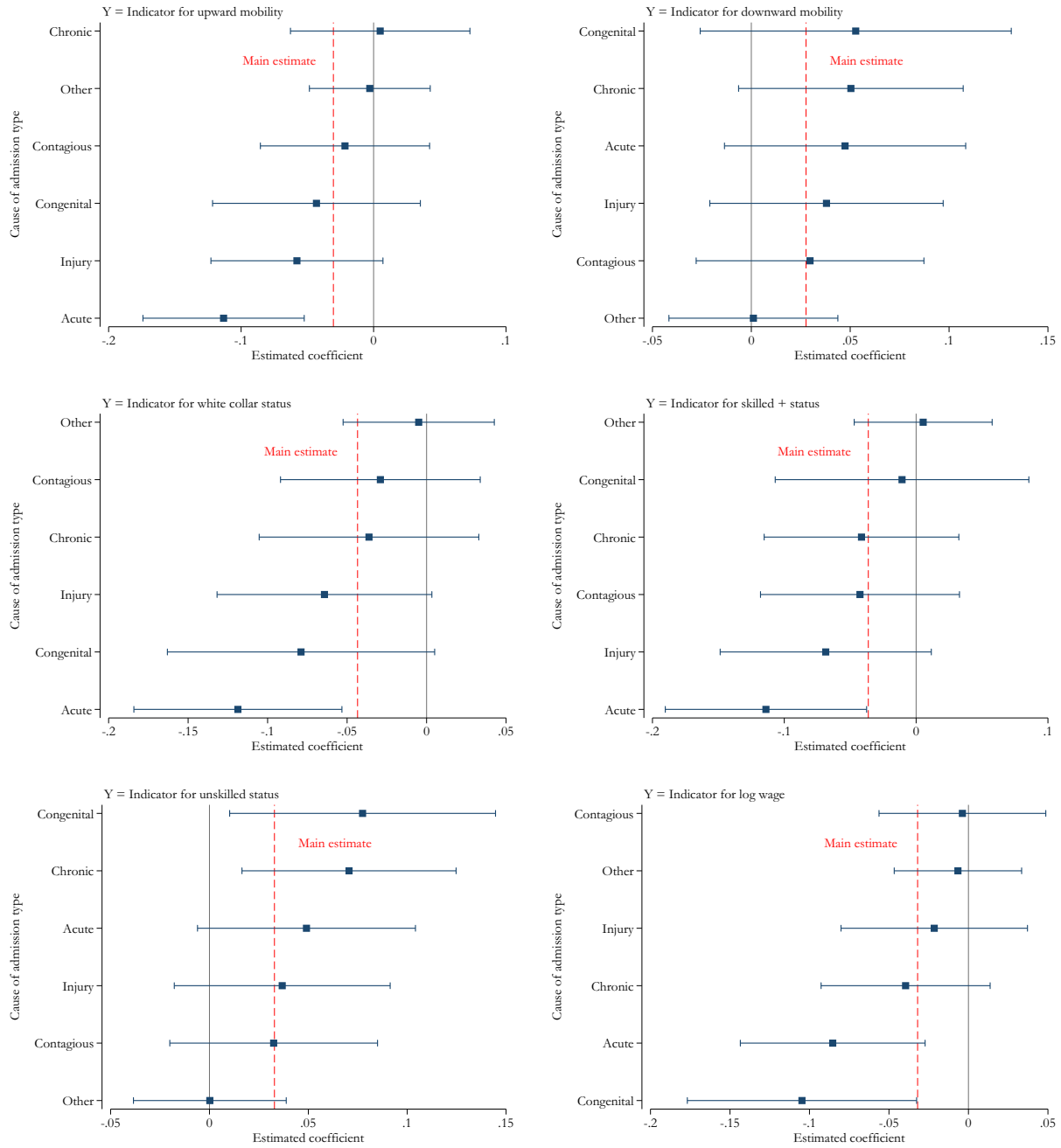


Figure A9: Randomly assign lower father's SES to 15% of patients 1000 times



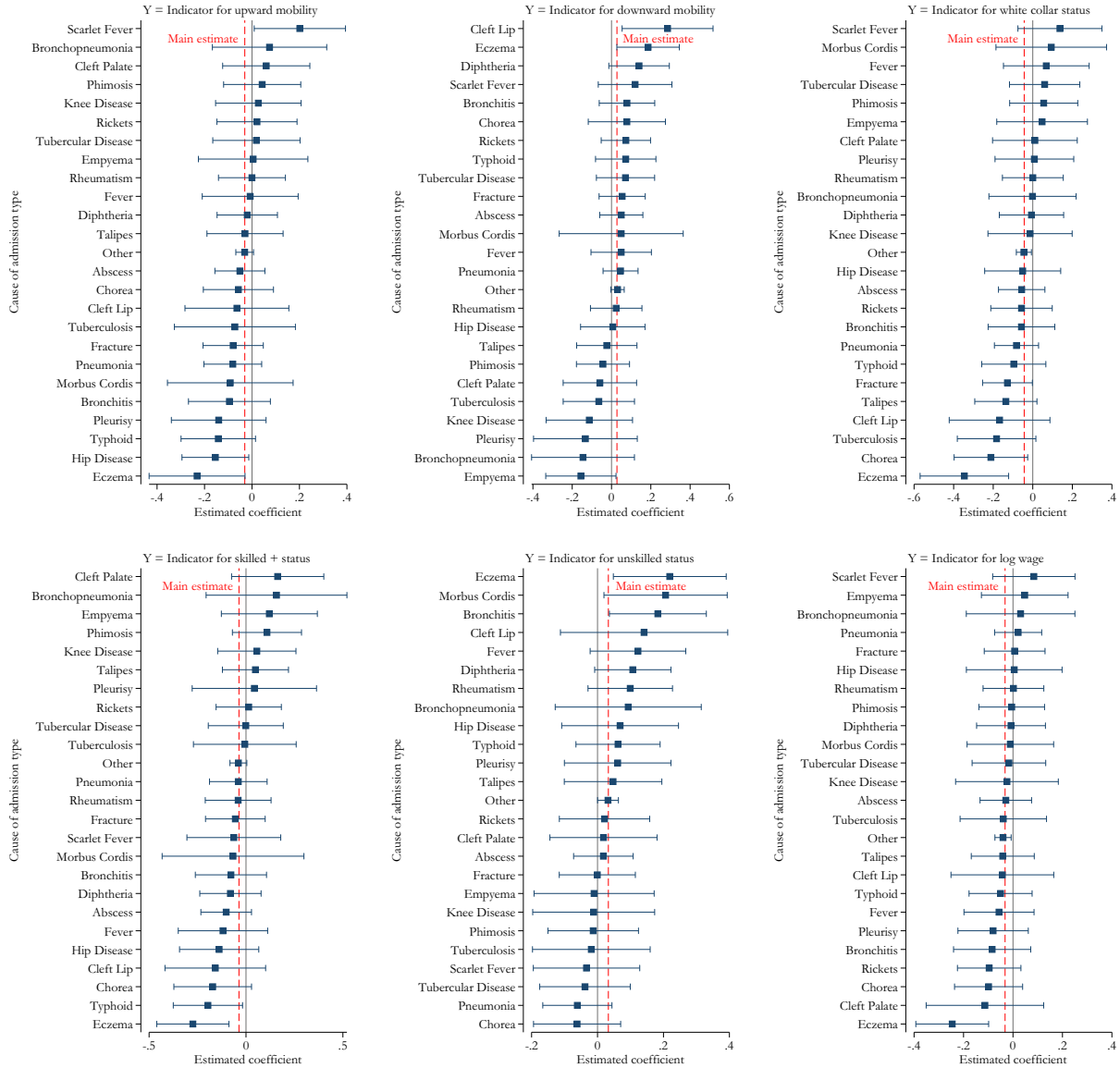
Notes: Each figure plots the distribution of estimated coefficients on the patient indicator from 1000 iterations of assigning the father's socioeconomic status (SES) to be one rank lower for 15 percent of the patients in the sample and re-estimating the main specification for the six dependent variables shown in Tables 2. When the father's SES is the lowest rank, it is assigned to be one rank higher for the non-hospitalized sibling only.

Figure A10: Effects of hospitalization by cause of admission type



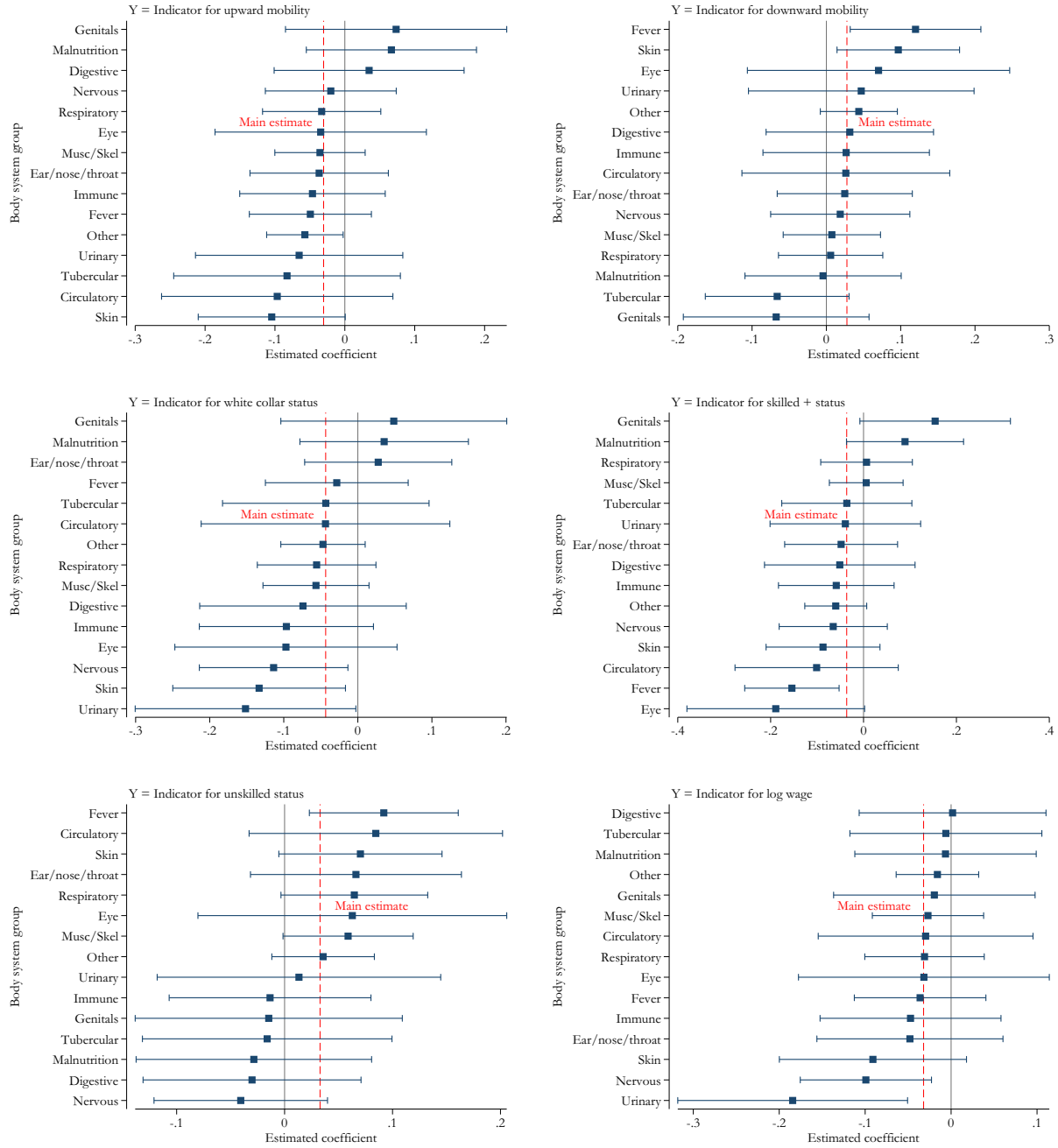
Notes: Each figure plots the estimated coefficients on indicators for a type of hospital admission and 95-percent confidence intervals from a single sibling fixed effects regression with one of six long-run occupational outcome variables shown in Table 2. In addition to the causes of admission plotted, the regressions include an indicator for admissions for multiple categories, as well as the same set of control variables listed in Table 2. Standard errors are clustered by childhood household.

Figure A11: Effects of hospitalization by common causes of admission



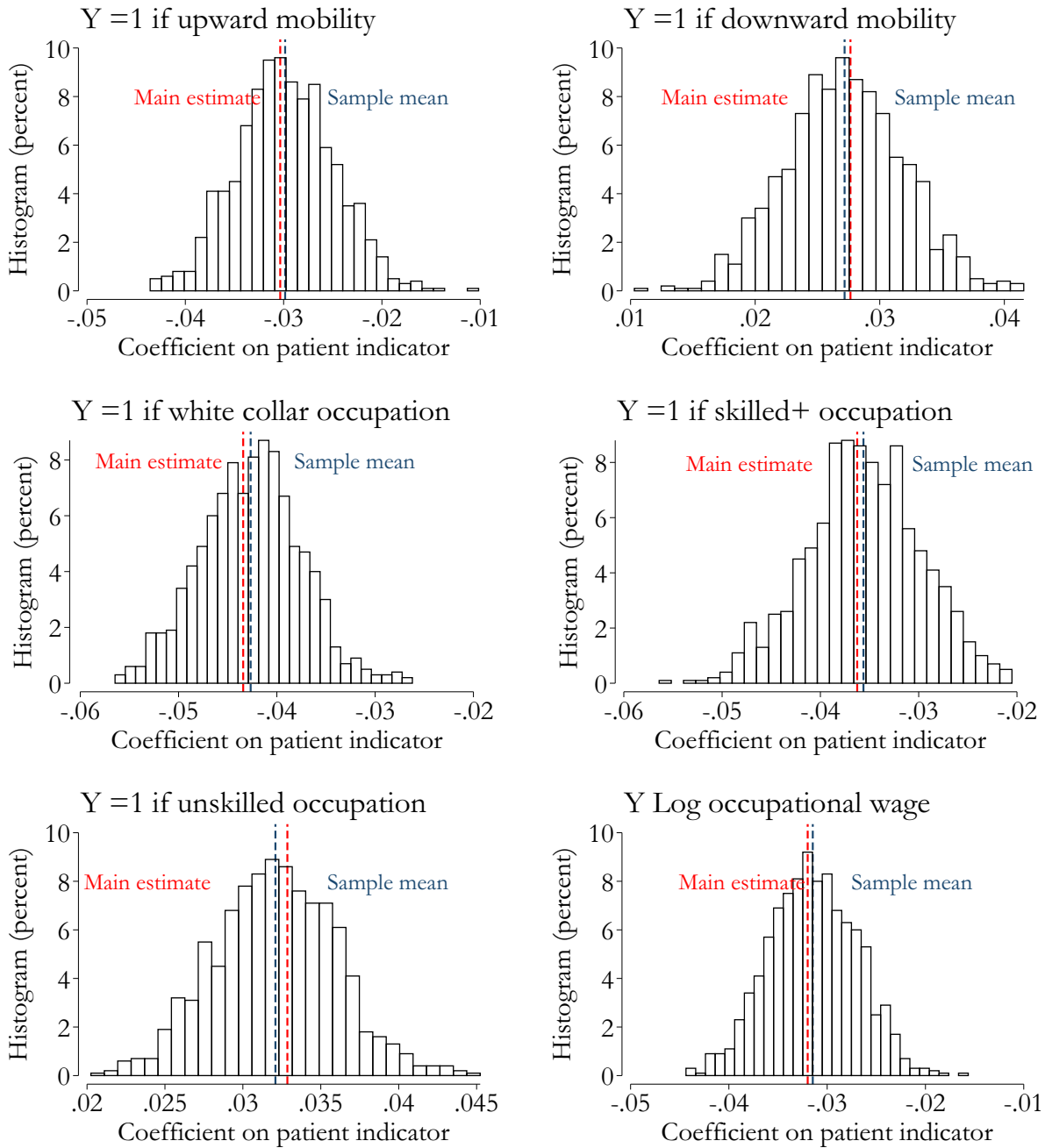
Notes: Each figure plots estimated coefficients, on indicators for the most common causes of admission listed in Table A2, and 95-percent confidence intervals from a single sibling fixed effects regression with one of six long-run occupational outcome variables shown in Table 2. In addition to the causes of admission plotted, the regressions include an indicator for admissions for multiple categories, as well as the same set of control variables listed in Table 2. Standard errors are clustered by childhood household.

Figure A12: Effects of hospitalization by body system classification



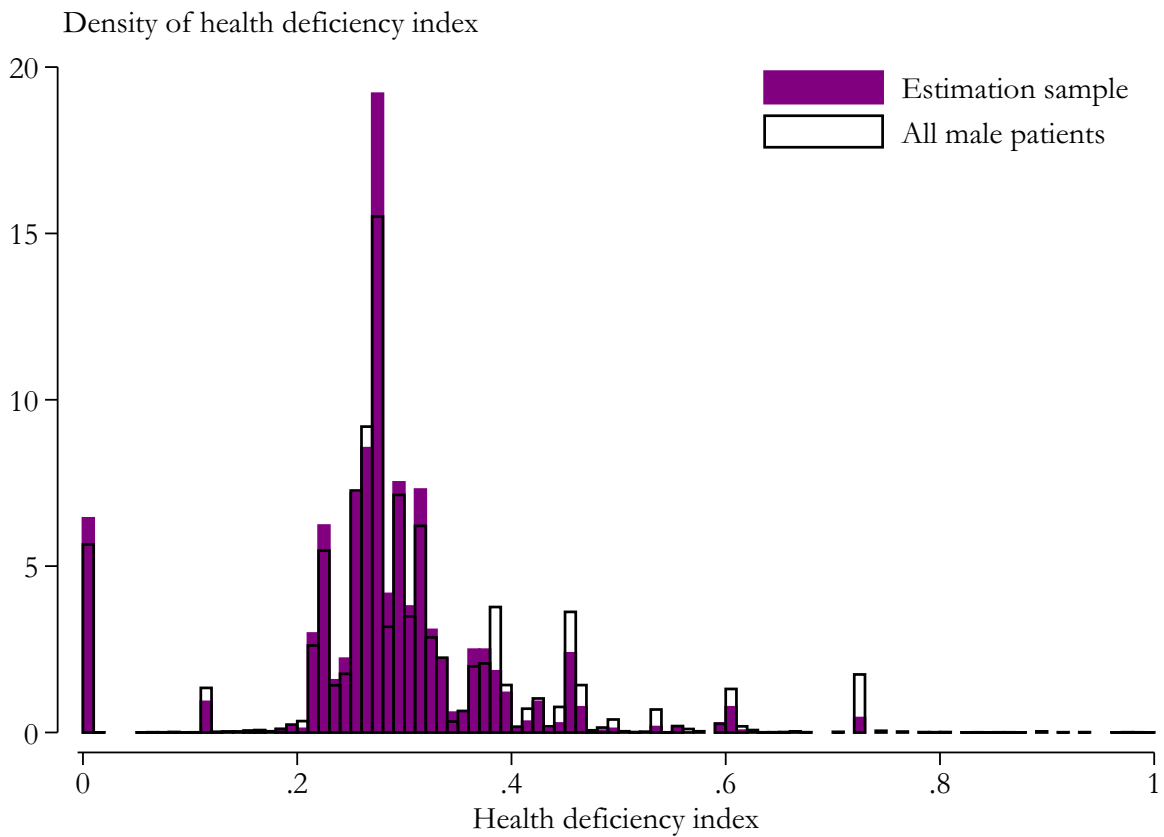
Notes: Each figure plots the estimated coefficients on indicators for the type of hospital admission, classified by body system, and 95-percent confidence intervals from a single sibling fixed effects regression with one of six long-run occupational outcome variables shown in Table 2. In addition to the causes of admission plotted, the regressions include an indicator for admissions for multiple categories, as well as the same set of control variables listed in Table 2. Standard errors are clustered by childhood household.

Figure A13: Randomly assign treatment (=1) to 10% of siblings 1000 times



Notes: Each figure plots the distribution of estimated coefficients on the hospitalization indicator from 1,000 iterations of assigning the treatment indicator equal to one for 10 percent of the non-hospitalized siblings in the sample and re-estimating the main specification for the six dependent variables shown in Table 2.

Figure A14: 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). See appendix B for a description of the procedure used to construct the health deficiency index.