Childhood health and long-run economic opportunity in Victorian England*

Krzysztof Karbownik† Anthony Wray‡

October 27, 2015

Abstract

We study the long-run effects of childhood health deficiencies on occupational outcomes for males, in addition to marital status for both genders, using a sibling-fixed-effects identification strategy. We construct a longitudinal data set consisting of in-patient admission registers, from two hospital in London, England, during the late-nineteenth century, linked to census and marriage records. Health deficiencies during childhood, proxied by the in-hospital mortality rate of the admitted condition, decrease the occupational log wage, and increase the probability of downward mobility relative to one's father, for male patients in comparison to their brothers. The effect of health deficiencies offsets 30 percent of the effect of an increase in father's occupational log wage on own occupational log wage. Female patients were less likely to be married as adults in comparison to their sisters, but there is no evidence of a marriage penalty for males.

*We are grateful to Susan 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 Barts hospital records. Wray is indebted to his dissertation committee members Joel Mokyr (chair), Joseph Ferrie and David Dranove for encouragement and guidance. We also thank David N. Figlio, Guillermo Marshall, Werner Troesken, seminar participants at Northwestern University, Hitotsubashi University, the Max Planck Institute for Demographic Research, and the Federal Reserve Bank of Chicago, conference participants at the World Congress of Cliometrics (Hawaii, 2013), the European Historical Economics Society (London, 2013), the Economic History Association (Washington, DC, 2013), the Social Science History Association (Chicago, IL, 2013), and the NBER Children's Meeting (Cambridge, 2015) for helpful comments and suggestions. Wray appreciates the financial support from the Northwestern University Economics Department's Eisner Fund and Center for Economic History, and the Economic History Association's Sokoloff Fellowship. All errors of omission are our own.

†Institute for Policy Research, Northwestern University; Email: krzysztof.karbownik@northwestern.edu.

‡Hitotsubashi Institute for Advanced Study, Hitotsubashi University; Email: anthony.wray@r.hit-u.ac.jp.

1 Introduction

An extensive body of research has shown that health and investments during the *in utero* and early-childhood periods have persistent effects on human capital accumulation and adult socioeconomic outcomes.[1] The effect of health deficiencies can work through not only biological changes, but also the inability to attend school, enter into an apprenticeship, or perform physically demanding tasks (Horrell *et al.*, 2001). Childhood illness, in particular, may weaken resistance to secondary infections, causing poor health to persist into adulthood.

Though poor childhood health has detrimental effects on labor market outcomes in the developed world today, the extent to which this phenomenon generalizes to other health environments or time periods is unclear. Using newly collected data on admissions to hospitals in late-nineteenth century London, England, this paper presents evidence that poor childhood health has a long-run scarring effect on crucial socioeconomic outcomes. We view hospitalization as a proxy for the totality of health deficiencies from the in-utero period through childhood, and examine its effects on individual occupational success, intergenerational mobility, and marital status for male patients. We also add to the literature's understanding of the effects of poor childhood health on women's socioeconomic status by constructing a unique panel of female patients linked to households in adulthood.

We focus on children hospitalized in London during the late-nineteenth century, a period characterized by significant health risks, as well as the limited availability of medical treatment and the lack of effective mitigation strategies.[2] Our approach makes three contri-

[1] This literature originated from efforts to evaluate the fetal origins hypothesis (Barker, 1990), which suggested an association between the intrauterine environment and the onset of chronic diseases in old age. It has since expanded to address a broader concern about "early influences" or "developmental origins" that spans the *in utero* environment, neonatal period, infancy, early childhood, and even adolescence (Gluckman and Hanson, 2006; Heckman, 2007). Economic evidence on the importance of the prenatal environment has come from studies of twins, siblings and singletons, and has documented significant long-run consequences of poor neonatal health commonly proxied by low birth weight (Black *et al.*, 2007b; Figlio *et al.*, 2014; Oreopoulos *et al.*, 2008; Royer, 2009). See Almond and Currie (2011a) and Almond and Currie (2011b) for extensive reviews of the fetal origins and early-childhood health literatures.

[2] During the nineteenth century, childhood illness could often be attributed to the overcrowded housing and poor sanitary conditions of resource constrained working-class neighborhoods (Szreter, 2005; Wohl, 1977, 1983), in addition to poor nutrition, conditions that are still present in the modern developing world (Fogel, 2004; Floud *et al.*, 2011). Given that the scientific elite did not reach a consensus on the germ theory of

butions to the existing literature on the long-run consequences of childhood health. First, we show that a broad range of health conditions, which have a long exposure window or insult individuals repeatedly, can affect later-life outcomes, in contrast to the rare events that the historical studies have relied on for identification.[3] Second, by turning to historical records we can draw inferences from a disease environment arguably similar to those found in some developing countries today, notwithstanding differences in public health infrastructure and the availability of antibiotics. The lack of census and administrative health data in many parts of the developing world today limits the scope of the analysis that researchers can conduct.[4] Moreover, present day hospitalization confounds the negative effects of health insults and the positive effect of hospital treatment, but historical hospital admissions can more cleanly identify the effects of poor underlying health in the absence of effective treatments. Third, we address the open empirical question regarding what factors explain the intergenerational persistence of poverty and downward social mobility, and suggest that poor child health was a contributing factor in late-nineteenth century England. Since we also observe the occupation of an individual's father, we can compute an intergenerational occupational elasticity and quantify what fraction of this estimate is offset by the effects of poor childhood health.

We construct a longitudinal data set consisting of hospital admission registers linked to

disease transmission until the 1880s, society had imperfect knowledge of preventative health behaviors such as basic hygiene practices (Mokyr, 2000; Worboys, 2000).

[3]Previous literature has studied sharply timed shocks like influenza pandemics (Almond, 2006; Brown and Thomas, 2013; Kelly, 2011; Lin and Liu, 2014), large-scale organized interventions such as the eradication of tropical diseases (Bleakley, 2007, 2010; Hong, 2007, 2011; Venkataramani, 2012), early-childhood disease exposure (Zhang, 2014), and the introduction of antibiotics (Bhalotra and Venkataramani, 2012; Jayachandran *et al.*, 2010), or used modern individual-level hospital admissions data, but only to explore effects on medium-run outcomes like schooling (Currie *et al.*, 2010). These studies found either negative effects of events detrimental to health or positive effects of public health interventions. Similar evidence has been shown for the developing world by Miguel and Kremer (2004), who evaluate the school-based random assignment of deworming drugs to Kenyan school children and find that the deworming treatment increased school attendance in the first year. Although they do not find effects on test scores, a follow-up study finds that children who participated in the deworming program experienced increases in hours worked and wages between the ages of 19 and 26 (Baird *et al.*, 2012). For a review of the literature on childhood health and adult outcomes in developing countries, see Currie and Vogl (2013).

[4]As was the case historically, serious and unexpected illnesses remain a significant threat to the economic opportunities of households in low-income countries, especially in the absence of health and disability insurance (Gertler and Gruber, 2002).

census and marriage records, in which it is possible to observe an individual's health status during childhood, in addition to occupational and marriage market outcomes during adulthood. In the absence of data on individual wages and educational attainment, occupational class and occupational log wages, as well as marital status, represent the most informative measures of socioeconomic opportunity that were recorded in historical documents for the population of Victorian England. A sibling fixed effects identification strategy is used to compare hospital patients to siblings of the same gender, who lived in the same household during childhood and did not appear in the hospital registers.[5] Thus, our estimates control for environmental factors common to the childhood household, as well as partially for unobserved genetic and health traits. However, a limitation of the sibling fixed effects model is that it fails to account for the reinforcing behavior of the parents, and thus we discuss this possibility at length in Section 4.2.

The sibling fixed effects models indicate that poor health in childhood had detrimental effects on the occupational class, occupational log wage, and prospects for intergenerational mobility of male patients as adults in comparison to their brothers. Our preferred estimates imply that a one-standard-deviation (9.7 percentage point) increase in the in-hospital mortality rate for the admitted condition, an index that we use as a proxy for the severity of the health deficiency, decreases the probability of attaining a white collar occupation by 11.4 percent, and the probability of entering a white collar or skilled occupation by 4.9 percent. It also lowers occupational wages by 0.068 standard deviations (3.3 percent) and increases the probability of downward mobility – attaining a lower occupational status than one's father – by 0.054 standard deviations (2.5 percent). These results are consistent across different specifications of the treatment and outcomes. They are invariant to different matching algorithms, bounds on spacing between patients and their siblings, as well as sample restrictions that exclude patients admitted multiple times, as infants, or with contagious diseases. On

[5]Previous studies that have used a sibling fixed effects identification strategy to address similar questions include Currie *et al.* (2010), Parman (2013), and Smith (2009), which we discuss below. Other studies that use linked historical census data and sibling fixed effects are limited to Abramitzky *et al.* (2012) who study migration and selection, and Mill and Stein (2012) who study race and labor market discrimination.

the other hand, the effects are larger for more severe diseases, for children hospitalized at older ages, and for those admitted to the general hospital instead of the children's hospital. Although we do not find any effects on the probability of marriage for males, we do observe a large and statistically significant penalty on the likelihood of being married for females with poor childhood health in comparison to their healthier sisters.

In order to fix ideas, consider the fact that the mean of the mortality index for all hospitalized patients was 10.6 percent, while a standard deviation change in the index was 11.2 percent. Thus, the average effect of health deficiencies essentially meant a shift from suffering conditions (with a 10 percent mortality rate) such as arthritis, abscesses, rickets, a common fever, or a puncture wound, to conditions (with a 20 percent mortality rate) such as nephritis (a complication of scarlet fever), heart disease, diphtheric paralysis, pain in the stomach, or a perforating wound. We can also attach an economic meaning to the magnitude of these effects by scaling the coefficients relative to the intergenerational wage elasticity, which we estimate by regressing the occupational log wage of fathers on that of sons. We find that a one-standard-deviation increase in the mortality index offsets 30.3 percent of the effect of a one-standard-deviation increase in a father's occupational log wage.

Our results suggest that individuals who suffered from poor childhood health were likely to attain a low occupational status, both unconditionally and in comparison to their fathers, and thus experience a lower quality of life. They also earned lower wages as adults, which could have lead to greater dependence on family and state support in old age. Finally, our results point to the need for a social safety net in economies where children have a high probability of experiencing health deficiencies. Although quantifying the 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.

Women have often been ignored in historical data because name changes at marriage

prevent the linkage of their census data from childhood to adulthood.[6] By matching females from childhood census records to Church of England marriage certificates, which contain both the maiden name and the married name, as well as the father's name, we can link information about women from childhood to census records during adulthood. We use marital status as a measure of economic opportunity for women at the turn-of-the-twentieth century since marriage insured women against poverty by providing access to a share of income from multiple wage earners in a household. Our results indicate that hospitalization during childhood was associated with a 8.6 percentage point (26 percent) lower probability of marriage for female patients relative to their sisters and thus potentially higher rates of poverty in adulthood.

The results in this paper complement Currie *et al.* (2010), who use a sibling fixed effects research design to show that poor health of children hospitalized in modern times is associated with greater welfare participation in young adulthood, primarily because it predicts poor health later in adolescence. We extend these findings by tracking individuals over the course of their working careers, up to the age of thirty-seven, and measuring occupational and marital outcomes, variables which Currie *et al.* (2010) do not observe in their data. Our data also contain a wide range of diseases in contrast to the focus on admissions for asthma and attention deficit hyperactivity disorder in Currie *et al.* (2010). Our results also complement Case and Paxson (2008, 2010) who use height as a proxy for childhood health and find that it is positively associated with cognitive test scores and levels of educational attainment.[7] Height captures the net influence of health throughout childhood and cannot evaluate the relative importance of shocks at different ages, whereas the hospital data can identify the age at which insults to childhood health occurred. The measures of health constructed from the historical hospital data are also more informative and reliable than the

[6] An exception is Olivetti and Paserman (2013), who construct a pseudo-panel of fathers and children of both sexes grouped by first name, and find an increase in intergenerational income elasticities between 1870 and 1930 in the U.S.

[7] Parman (2013) also uses height as a proxy for childhood health and compares outcomes for brothers in the World War II U.S. military enlistment records to show that an increase in height at enlistment corresponds to a small but significant increase in educational attainment.

self-reported health measures used in many contemporary studies (Smith, 2009).

The remainder of the paper proceeds as follows: Section 2 provides historical background on children's hospitals and health care institutions in London; Section 3 describes the data, outlines the linkage procedure, and provides a descriptive analysis of the hospital admissions data; Section 4 explains the estimation strategy; Section 5 describes the main results; Section 6 presents robustness checks; Section 7 examines heterogeneity in the effects by admission characteristics; Section 8 provides background on Victorian- and Edwardian-era marriage markets and analyzes effects on female patients; and Section 9 concludes.

2 Historical Background

2.1 Childhood health and hospital care before the 19th century

Prior to the establishment of children's wards in general hospitals and the advent of specialized children's hospitals in the middle of the nineteenth century, children had limited options for receiving formal medical care in England. The earliest forms of institutionalized care for children were limited to foundling hospitals, nurseries in general hospitals, and orphanages, which essentially served the poor and needy, and provided little medical care. The Foundling Hospital in London opened in 1741 and admitted all infants that appeared at its doorstep, a policy which contributed to a 70 percent mortality rate for admissions between 1756 and 1860 and perpetuated the view among doctors that infants should not be separated from their mothers (Franklin, 1964, p. 104). Until the mid-nineteenth century, mothers opposed the admission of infants to hospitals on the basis of the belief that hospitalization undermined parental authority and that the baby would suffer ill-effects from the separation from her mother.[8]

There were few exceptions to the rule of excluding child patients from general hospitals.

[8]Efforts to convince the public about the benefits of hospitalization were not helped by a lack of consensus among medical practitioners over whether children received better care at home or in a hospital (Strange, 2005, p. 47).

Guy's Hospital in London admitted children to its women's wards from its inauguration in 1722, but did not establish a children's ward until 1866.[9] St. Thomas' Hospital admitted only 129 children between 1773 to 1796, accounting for no more than 5 percent of admissions.[10] Overall, in London, hospital admissions by children under the age of ten accounted for only 1 percent of hospital admissions (26 of 2,363 patients) in 1843, despite the fact that deaths by children represented half of total mortality (Higgins, 1952, p. 10). During the middle of the nineteenth century, general hospitals in London admitted more sick children as inpatients, particularly those requiring surgical procedures or those suffering from fevers. London Hospital opened a children's ward in 1840 with eighteen beds, but by 1857, only treated 214 patients under the age of seven, while St. Bartholomew's admitted children in regular wards (Lomax, 1996, p. 20).

Before the nineteenth century, neither the state nor the medical profession viewed the long-term survival of children as its responsibility. Instead, responsibility for child care fell to charities in the absence of parental support. The general tolerance for high rates of child mortality and perception of children as poor and weak changed with the European Enlightenment as people began to view the survival of infants and healthy child development as vital inputs to the continued existence and growth of society (Seidler, 1989, pp. 181-3). Investment in child health promised not only returns for parents in the form of child labor, but also an opportunity for doctors to increase medical knowledge by admitting "interesting" medical cases (Levene, 2012, pp. 122-3).

The origins of children's hospitals can be traced to a shift in the perception of hospitals during the eighteenth century from places of refuge for the poor and destitute, to institutions (*Krankenhaus*) for the treatment of sick patients (Seidler, 1989, p. 181). The system of institutionalized care for children that developed in London during the late-eighteenth and early-nineteenth centuries lagged behind continental Europe as it only provided outpatient

[9]However, between 1831 and 1850, Guy's Hospital built fifteen "cribs" for children in a wooden building constructed over stables (Higgins, 1952, p. 10).

[10]During this period, the admission totals included both inpatients and outpatients, but not all outpatient visits were recorded (Levene, 2012, p. 122, 147).

medical services to the children of the working poor. The first outpatient dispensary opened in 1769 and treated 35,000 children (3,000 per year) before closing in 1782 (Seidler, 1989, p. 185), followed by the Royal Universal Dispensary for Children that admitted 175,000 children (5,800 per year) between 1816 and 1846 (Franklin, 1964, p. 109). Meanwhile, the Hopital des Enfants Malades in Paris was founded in 1802 as the world's first children's hospital, with a capacity of 250 beds for the treatment of acutely ill patients. London would remain without a specialized children's hospital until Great Ormond Street Hospital (GOSH) was founded by Charles West in 1852 (Seidler, 1989, p. 185). The children's hospitals in England also trailed the French hospitals in terms of accessibility as they were established with the intention of providing free care to the "deserving poor" – the class of patients above pauperism, but below the capacity of paying for medical care (Lomax, 1996, p. 41) – whereas the French sought to provide medical services for all citizens (Seidler, 1989, p. 183). Initially, GOSH had a capacity of ten beds and would grow to house sixty-two beds by 1864, before expanding to 120 beds following the construction of a new building in 1877 (Franklin, 1964, p. 112). Soon, other children's hospitals opened in London, with five hospitals operating in 1869 and eleven by 1890.

2.2 Health care institutions

London's hospital system consisted of decentralized and unregulated charitable institutions, known as voluntary hospitals, that received no state funding. The largest hospitals, St. Thomas', St. Barthlomew's and Guy's, financed operating costs out of large endowments, but other hospitals relied on a system of charity subscriptions and public appeals for donations that often tied the provision of medical services to a specific location. The growth of the hospital market did not follow the geographic expansion of London's population as the majority of the hospitals were located within two square miles of the metropolitan center (Ball and Sunderland, 2001, p. 368). Figure A1 plots the locations of general and children's hospitals in central London and points towards the limited access to health care facilities

in the poorer eastern districts. It also highlights the limitations with historical data, given that admission records have survived for only four of the ten largest general hospitals, and two of the five largest children's hospitals in London.[11]

A large and expanding network of health care institutions served the sick and destitute in London during the latter-half of the nineteenth century, including the voluntary hospitals which were categorized as general hospitals, teaching hospitals (associated with a medical school) or specialty hospitals. The market for institutionalized health care also included dispensaries, which were funded by charity and competed with hospital outpatient departments for poor patients, and with private doctors for wealthier patients who did not require inpatient admission. The Poor Law Amendment Act of 1834 funded workhouses that provided care to sick paupers and the destitute, who were rejected by the voluntary hospitals (Abel-Smith, 1964, p. 46). The Sanitary Act of 1866 provided authority for the creation of isolation hospitals that helped to contain the spread of infectious diseases by admitting fever and smallpox patients who had been refused admission by workhouses and voluntary hospitals. The Metropolitan Asylum Board (MAB) established a series of hospitals under the Metropolitan Poor Act of 1867 for the treatment of paupers, chronic conditions, and smallpox cases, but a revised Public Health Act in 1891 gave all Londoners access to free treatment at the MAB hospitals (Abel-Smith, 1964, pp. 119, 126).[12] By the 1880s, germ theory and bacteriology were accepted as the consensus view on disease, and public health officials focused on policies of notification, isolation, and disinfection to identify carriers of disease and present further contagion (Worboys, 2000, p. 236).

Although gains in knowledge about bacteriology and the practice of performing autopsies increased the ability of physicians to identify the cause of illnesses, medical care remained

[11] In Section 3 we discuss issues that may arises from the selection of hospitals in our sample.

[12] In the 1870s, children's hospitals also began providing isolation facilities for patients with infectious diseases which limited the spread of infection within the hospital. In particular, the Great Ormond Street Hospital for Sick Children developed formal procedures that restricted the movement of medical staff between the isolation ward and the rest of the hospital as nurses who had treated infectious fever patients were required to wait outside for two hours before they could return to the main hospital wards (Tanner and Hawkins, 2013, p. 219).

limited to the treatment of symptoms and the provision of therapeutic benefits prior to the introduction of antibiotics. In the majority of cases, the medical treatment of children consisted primarily of bed rest, nursing care, and the provision of an adequate diet, with the aim of recovering strength and increasing resistance to disease. Especially among poorer patients, children also benefited from a more sanitary environment than the crowded conditions at home. More notable gains in hospital practices were made in surgery. Prior to the introduction of anesthetics, surgical procedures for children were restricted to the lancing of abscesses, resetting broken bones, amputation of limbs, and the removal of bladder stones (Lomax, 1996, pp. 98, 103, 124). From the middle of the nineteenth century, improved knowledge about the causes and control of infections, as well as the increasing use of anesthesia, expanded the scope of surgical procedures that could be performed in hospitals. Major operations included the removal of tubercular glands or the appendix, as well as the repair of congenital malformations such as cleft palate or club foot. While the number of operations performed at St. Bartholomew's Hospital increased from 417 in 1863 to 2,446 in 1899, the mortality rate of surgical procedures fell by half to 7 percent over the last quarter of the nineteenth century (Medvei and Thornton, 1974, p. 219).

Although the hospitals could offer little in the way of treatment for illness, the admissions data suggest that improvements in sanitary practices and the isolation of patients with infectious disease contributed to improvements in the quality of hospitals. Figure A2 plots the coefficients on year-of-admission dummy variables in a regression of an indicator for an admission resulting in death in the hospital on patient and admission characteristics. The plot shows that after adjusting for patient characteristics and the admissions practices of the hospitals, the probability of death in the hospital trended downward over the second-half of the nineteenth century. This finding suggests that the underlying quality of the hospitals had improved.

Hospitals in England continued to evolve during the second half of the nineteenth century, resulting in changes in the selection of patients admitted. During the 1860s and 1870s, the

patients of voluntary general hospitals were drawn primarily from the London poor and working class. The original mandate of voluntary hospitals was to provide palliative care for the sick poor, a target population that was distinct from the destitute. Hospitals sought to deny admission to patients in receipt of public aid under the Poor Law, though they could not always distinguish between the needy poor and those on relief. Hospitals recognized that the wealthy had no interest in entering an institution, but middle class patients were under-served by hospital care. Furthermore, the growing knowledge of the infection process made it clear that hospitals ought to provide care to all classes of society (Abel-Smith, 1964, p. 119). Eventually, voluntary hospitals developed into institutions where doctors treated middle- and upper-class patients, trained medical students, and gained authority as sources of new medical knowledge. Meanwhile, the privileged upper-classes of England continued to rely on general practitioners who operated private clinics and treated patients in the privacy of their homes, even though medical education focused on hospital care and doctors could not make use of new medical technologies such as x-ray machines in the context of home care (Carpenter, 2010, p. 25).

Acceptance of hospitals as safe venues for medical care was not universal as some working-class families continued to associate hospitals with pauper workhouses and venues of death, and refused to seek professional medical advice during the early stages of illness.[13] Instead, families faced the risk of contagion and nursed the sick and dying at home with the assistance of informal mutual aid networks among relatives and neighbors (Lomax, 1996; Ross, 1993; Strange, 2005). Other Londoners refused to comply with public health regulations and considered medical procedures such as injections and surgery to be "unnatural and dangerous assaults" on the body.[14] Even measles was not considered a potentially life-threatening illness, but rather a childhood rite of passage (Strange, 2005, p. 40).

[13] In 1871, an observer in Sheffield remarked that "it is quite common for women to defer sending for medical aid, when the children are ill until it is too late" (Wohl, 1983, p. 18).

[14] And yet, self-remedies included tying a piece of bacon around a girl's neck for three weeks to relieve a sore throat, and taking children to the river at low tide to breathe in the sulphurous fumes (Ross, 1993, p. 174, 177).

2.3 Hospital admission practices

As hospital budgets fell into deficit in the 1870s, a belief emerged among hospital staff that patients who could afford the fees for private medical service were abusing the hospital system (Lomax, 1996, p. 9). Efforts by hospital management to identify and restrict access of patients who could afford the payment of admission fees proved to be unpopular and, ultimately, unenforceable. London general hospitals began accepting payments from patients in the 1870s as the cost of treating inpatients rose and fund-raising sources dried up. Thus, while voluntary hospitals were originally conceived as facilities to care for the working poor, they admitted patients from different socioeconomic groups during our period of study, which covers admissions between 1874 and 1901.

The process of admission to a hospital began with the would-be patient's mother bringing her child to the hospital's outpatient department. Upon entering the hospital after an extended wait in line, the parent was screened by a clerk to determine her ability to pay for medical treatment, while the child was examined by the house surgeon or physician on duty to determine suitability for admission as an inpatient. Children who were not admitted still received some medical advice as the hospital hoped to spread knowledge of health behaviors to the working class (Tanner, 2007; Tanner and Hawkins, 2013). At the end of our sample period in 1900, the children's hospital admitted over 1,600 inpatients, accounting for approximately 8 percent of the 20,000 children examined in the outpatient department (Tanner and Hawkins, 2013, p. 217).

In practice, doctors had considerable authority over the types of cases that were admitted from the pool of outpatients that far exceeded a limited supply of hospital beds. Medical staff would selectively admit patients from the outpatient department, favoring medically interesting cases and acutely sick patients in order to demonstrate immediate results of treatment for teaching purposes (Abel-Smith, 1964, p. 39). Doctors were opposed to admitting chronic or incurable cases in order to avoid having a bed occupied for a lengthy period of time, or to limit the number of deaths at the hospitals (Waddington, 2000, p. 9). The more

patients a hospital treated and the fewer deaths reported, the more attractive the hospital appeared to donors when it applied for funds. Thus, the set of diseases admitted to hospitals are not necessarily representative of the population.[15]

3 Data

We construct a linked longitudinal sample that combines individual-level information from three primary sources: hospital admission registers, population census records, and Church of England marriage certificates from London parishes. We use the causes of admission contained in the hospital records as a proxy for poor childhood health. The population census records from 1881, 1891 and 1901 provide us with information on the family structure during childhood, while censuses in 1901 and 1911 give insight into labor market outcomes during adulthood. Marriage certificates document the name changes needed to track females from childhood to adulthood. Given that it is not possible to examine effects on occupational outcomes of married females, who rarely worked, we analyze male patients and their brothers in one sample, as well as female patients and their sisters in another. First we present the results for labor market outcomes and intergenerational mobility based solely on the sample of males before turning to a discussion of the marital outcomes for both genders in Section 8.

3.1 Hospital admission registers

We create a new data set using the individual-level inpatient admission registers of two hospitals in London, England from the nineteenth century. We transcribed the records of male inpatient admissions to St. Bartholomew's Hospital (Barts), one of four large general hospitals in London, and obtained the admission records for patients of both genders who were admitted to the Hospital for Sick Children at Great Ormond Street (GOSH), the largest

[15] We present descriptive statistics in Section 3 that address the selection of patients into hospitals.

London-area children's hospital (Kingston University, 2010).[16] The hospitals admitted a majority of their patients from neighboring districts (Figure 1) and 83 percent of patients from the County of London. For both hospitals, we collect information on the universe of admissions of male patients from the 1874 to 1890 birth cohorts who were hospitalized between the ages of 0 and 11 years. As shown in Figure 2, children under the age of 13 did not participate in the labor force, so we do not need to worry about hospital admissions related to child employment. For female patients we have only transcribed the records from GOSH. An individual entry in the hospital admission records typically includes the patient's name, age in years, and residential address; a description of the patient's cause of admission; dates of admission and discharge; and the name of the attending physician, which we use to determine whether the patient was treated by a doctor.[17]

In the absence of unique and permanent patient identifiers in nineteenth-century hospital admission records, we construct a patient identifier based on unique combinations of first name, surname, year of birth, and district of residence. Among admission records matched to the census, we redefine patient identifiers when multiple admissions are uniquely matched to the same individual in the census.[18] We observe 5.6 percent of patients admitted multiple times in the transcribed sample, accounting for 11.7 percent of admissions by children aged 0 to 11. In the main specification, the hospital treatment takes on the value one if the patient ever satisfies the condition of interest, or the maximum value across all admissions, but our main results are invariant to different methods of treating cases with multiple observations per individual.[19]

[16]These records were transcribed by volunteers in London from inpatient admission registers as part of the Historical Hospital Admissions Research Project (HHARP). As a rule, the children's hospitals did not admit patients above the age of 12 years, with some exceptions. We do not make use of an incomplete series of admissions to Evelina Children's Hospital, which do not affect the results if added to the sample.

[17]In the case of GOSH, we also observe a categorical variable for the outcome of the hospitalization (cured, relieved, not relieved, or died). Apart from identifying patients who died in the hospital, the hospitalization outcome variable is not very informative about the prospects for recovery. Hospitals tended to inflate the extent to which patients were cured or relieved in an effort to attract and maintain donors.

[18]See Appendix 9 for a detailed description of the process of constructing the patient identifiers.

[19]In Section 6.1 of the robustness analysis, Tables A12 and A9 present results using various measures of accounting for multiple admissions. In particular, we drop multiple admissions, restrict attention to the first observed admission, or take the sum of all admissions.

Even though we have data from only two hospitals in London, both were among the largest hospitals in their respective categories and accounted for a sizable fraction of the market for inpatient hospital care.[20] Table 1 presents descriptive statistics on the number of beds, inpatient admissions, and outpatient visits in 1894 for both hospitals in relation to the twelve largest general hospitals and six largest children's hospitals. St. Bartholomew's Hospital accounted for 12 percent of inpatient admissions, 23 percent of outpatient visits, and had a capacity of 675 beds, which corresponded to 20 percent of total beds among the twelve largest hospitals in London. Great Ormond Street was the largest and oldest children's hospital in England with a capacity of 178 beds in 1894, and covered a large share of the children's hospital market with 29 percent of inpatients, 25 percent of outpatients, and 36 percent of beds.

3.2 Data linkage

We implement an iterative strategy used prominently in the literature to determine whether individual records are successfully matched across databases (Abramitzky *et al.*, 2012; Aizer *et al.*, 2013; Long and Ferrie, 2013). Figure 3 illustrates the data sources involved in constructing the data set and the variables used to link the sources.[21] The first linkage is formed by locating the hospital patients during their childhood years in the 1881, 1891 or 1901 population census of England. We require the following variables to match between the hospital and census records: first name, middle initial (when reported), surname (with a degree of tolerance allowed for the similarity of text strings), and the year of birth ($\pm$ 1 year) implied by the ages reported in the records. We tolerate up to six other records enumerated in the census with similar names within 1 year of birth, and twelve records with similar names within 2 years of birth.[22] Given that census manuscripts are organized by

[20] We are currently in the process of digitizing records for both males and females of Guy's Hospital, the second largest general hospital in London at the time. We are also planning to add information on female patients admitted to Barts Hospital.

[21] See Appendix 9 for data references and a detailed description of the linking procedure.

[22] The number of individuals enumerated in the census born within one or two years of the matched record and with similar names provides a measure of the likelihood for a true match. A similar name is defined as

household of residence and include the relationship of each individual to the household head, it is straightforward to also collect the entries for all same-gender siblings of the hospital patients matched to the census. We retain all same-gender siblings and discard all households with either a singleton male child or one male child with only female siblings.

A second linkage is required to obtain outcomes as adults and involves identifying a unique match for the male patients and their siblings in the 1901 or 1911 censuses of England using name, year of birth, and birthplace. Individuals are matched between censuses either by entering the linking variables as search criteria, or by collecting census records from 1901 or 1911 that appear as "hints" to suggested records on the Ancestry.com web page displaying the childhood census record. We attempt to match all siblings to a census during adulthood and retain the matched sibling closest in age to the patient. If we fail to match either the patient or at least one sibling within 8 years of the patient's age we discard the household from the sample.[23] We prioritize records matched via the "hints" since some of these links are based on additional information not observable in the census. Any individuals without a unique match across all data sources are discarded from the sample.

A common limitation of linked historical census data is that the sources do not contain enough information to match all records uniquely. Panel A of Table 2 traces the attrition of observations in the sample at each stage of the linking procedure according to the restrictions on match quality that we impose. The two rows display match rates for Barts and GOSH, respectively. We begin with nearly 19,000 patients of whom we match 47 percent to a census during childhood.[24] This match rate is higher than what is typically found in census-to-census matching because we match individuals from hospital records to at least one of three censuses. Column (3) illustrates sample attrition due to the absence of male siblings in the same household. Since Victorian London was a society with relatively high fertility, only

a Levenshtein string distance between two name strings weighted by the length of the name string in the hospital records that takes a value of less than 0.1.

[23]Figures A3 and A4 shows that our results are not sensitive to varying the maximum age gap between patients and siblings.

[24]The figure reported for the total number of patients excludes patients who died in the hospital and patients who resided outside Greater London (the counties of London, Essex, Kent, Middlesex and Surrey).

about 12 percent of the initial sample is lost due to the restriction to same-gender siblings. In the next step (column 4), we experience large attrition as we require both the patient and at least one male siblings matched between censuses ten to thirty years apart. In column (5) we further refine the matched sample by removing observations with missing data and imposing restrictions to drop extreme outliers from the final sample.[25] The final sample used in the baseline estimation is comprised of 1,238 sibling pairs, each of which includes one patient from the hospital records and one sibling who was not observed in the hospital records.

Similar to other studies, our matched sample is not a random sample of the population. In particular, Table 3 shows that individuals in the final sample are more likely to be older at the time of admission to the hospital – which could be due to selective mortality in early childhood. However, we are less likely to match children with either very rare or very common names, as shown in Figure 4. The U-shaped relationship is due to difficulty matching names with atypical spelling variations and typographical errors in the transcription. Figure 5 shows that we are less likely to match patients admitted further away from census years, which could reflect selective migration. Our final sample is affected by selective mortality since we are also less likely to match those who were admitted with a high-mortality condition and those who were treated by a doctor, which proxies for a severe cause of admission.

Panel B of Table 2 shows the share of hospital records that do not match to a census taken during the patient's childhood, match uniquely, or match to multiple census records. Match failure can be attributed to mortality between the hospital discharge and census enumeration dates, inaccurate information in the hospital or census records, under-enumeration, or migration outside the county of residence at the time of hospitalization.[26] The share of

[25] In particular, we exclude individuals enlisted in the military because their occupations are not recorded in the census. We also require individuals to be no more than 18 years old when enumerated in their childhood household to avoid including a selected group of individuals who live with parents at older ages. See Section 4.1 for a description of other sample restrictions related to match quality.

[26] In our main specification, we opt not to use additional information on the district of residence to resolve multiple matches as it would only increase our sample by 15 percent and it would bias our sample in favor of matching individuals with less geographic mobility. Our results are unchanged when we include the additional observations.

uniquely matched records is similar across hospitals, but a higher percentage of observations from GOSH are not matched to census records due in part to a lower rate of reporting middle initials. Although a larger share of records fail to match in the second stage of the matching procedure, due to mortality and international migration, we match more records uniquely because parish and county of birth are used as additional matching variables.

3.3 Representativeness of the matched sample

A comparison of the matched sample to the population of England indicates that the occupational classes of patients' fathers are not representative of the population. Occupational titles are assigned to one of four categories: unskilled, low-skilled, skilled, and white collar, the latter of which includes professional, managerial, and clerical occupations. Figure 6 shows the occupational distribution of fathers in the final sample of households with hospital patients and siblings relative to the population. Figures A5c and A5d compare the occupational distribution for the sample of male patients and siblings as adults to the overall population in 1901 and 1911, respectively. In general, the proportion of in-sample households in low-skilled and skilled occupations is higher than in the population, while the share of in-sample households with unskilled and white collar occupations is lower than the population. These patterns reflect the fact that wealthier, middle-class and upper-class households could afford to pay for the services of a doctor in the privacy of their home and were less likely to seek admission at a hospital, while voluntary hospitals were reluctant to admit paupers and the extremely poor classes, who could access lower quality Poor Law hospitals.

3.4 Occupational outcomes in census records

The occupational titles reported in the 1901 and 1911 population censuses of England provide a measure of social class for the hospital patients and their siblings as working adults. We also obtain the father's occupation from the census to which patients and siblings were

matched during childhood and use it to compute measures of intergenerational mobility. We compare both unconditional occupational attainment between siblings and occupational success conditional on their father's occupational status when they were children. The latter measure provides an indication of the extent to which poor childhood health hinders the intergenerational transmission of status.

The Integrated Census Microdata Project (I-CeM) has released complete count files for the population censuses, in which a Historical International Standard Classification of Occupations (HISCO) code has been assigned to each of the unique occupation strings (Leeuwen *et al.*, 2002; UK Data Archive, 2014). We rank the socioeconomic status of the occupational titles according to the HISCLASS scheme, which assigns each of the 16,000 HISCO occupation codes to one of twelve social classes (Leeuwen and Maas, 2011). The assignment of the HISCLASS category is based on the extent of supervision and skill level required by the occupation, whether the occupation is manual, and by the economic sector of the occupation. We adopt an occupational ranking based on HISCLASS that contains seven social classes and has been used by the previous literature (Abramitzky *et al.*, 2011, p. 126):

1. Managers and professionals
2. Clerical and sales personnel
3. Foremen and skilled workers
4. Farmers and fishermen
5. Lower skilled workers
6. Unskilled workers
7. Lower-skilled and unskilled farm workers

We refer to class 1 and 2 occupations as white collar (e.g. clerk), class 3 and 4 occupations as skilled (e.g. cabinet maker), class 5 occupations as low-skilled (e.g. house painter),

and class 6 and 7 occupations as unskilled (e.g. general laborer).[27] Table A1 lists the most common occupations in each of the four groups that we use to construct binary dependent variables in the analysis of unconditional occupational attainment. In the robustness analysis we construct indicators of occupational class using an alternative classification of occupational titles developed by Armstrong (1972) based on the Registrar General's 1921 and 1951 classification schemes.[28] As an alternative, more continuous, outcome measure we use estimates of occupational log wages constructed by Williamson (1980) for one of twenty-two occupational groups. Following Ferrie *et al.* (2012), we impute wages based on the occupation of the individual when available, and otherwise assign the average wage for the occupational class. The Williamson (1980) wage estimates have been criticized on the basis of the sources used to assign wages for solicitors and barristers, surgeons and doctors, as well as engineers. These concerns have limited significance for our study given that wealthy professionals were rarely admitted to hospitals in nineteenth-century London. In fact, we only exclude two patients with the occupational title of "lawyer" from our matched sample.

3.5 Causes of admission to the hospital

A unique feature of the hospital admission registers is the detailed information on the causes of admission, which provides us with ample information about the health status of children in Victorian London and how the diseases from which they suffered compare to those prevalent in contemporary environments.

The richness of the conditions that proxy for childhood health allows us to construct multiple treatment variables in order to assess how robust our are results to different measures of weak health during childhood. The most basic treatment variable is an indicator for whether we observe a child in the hospital records. Although it is clearly the case that tuberculosis or infections indicative of a weakened immune system could potentially affect adulthood

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

[28] Our results are unaffected. See Ferrie *et al.* (2012) for more details on Armstrong classification system.

outcomes, we would not expect a minor wound or even a fractured bone to have long-lasting effects. Thus, a simple indicator for hospitalization may not provide the most informative measure of health deficiencies relevant in the long run. The second treatment measure splits the causes of admission into cases that we categorize as "acute" or "non-acute". The group of acute conditions includes cases that were described as "acute" in the hospital registers at the time of admission, conditions described as acute in a modern definition (Farlex, 2014), as well as the sequela or immediate complications of such conditions.[29] Panels A and B of Table A2 list the most common conditions in the two categories for the full sample of admissions, in addition to the matched regression sample. The acute category includes cases of diphtheria, a prominent infectious disease during the nineteenth century, as well as cases described as sequela of diphtheria or diphtheric paralysis, a common complication of the disease. We also include cases such as tuberculosis or pneumonia which may have an acute phase followed by chronic complications.[30] Causes of admission in the non-acute category are primarily fractures and other injuries due to accidents, in addition to chronic conditions such as diseases of the hip or knee.[31]

To avoid concerns about the subjective classification of conditions, we also abstract entirely from the description of the cause of admission in the hospital registers and construct an alternative treatment that relies instead on information about the identity of the attending physician who admitted and treated the patient. We proxy for health deficiencies with an indicator for cases overseen by a doctor rather than a surgeon or assistant medical provider. This distinction is identified by whether "Dr" or "Mr" was recorded next to the name of the

[29] See a debate over the classification of historical diseases in Condrau and Worboys (2007), Mooney (2007), and Condrau and Worboys (2009)

[30] Children who suffered from infectious diseases such as measles and whooping cough faced a higher risk of contracting tuberculosis, and if they survived, could have been left with permanent disabilities. Inflammation of the lungs (bronchopneumonia) was a common complication of these infectious diseases that could cause a collapse of a lung, from which it was not possible to fully recover. Individuals could also develop an inflammation of the inner ear (otitis) or an infection of the cornea (ophthalmia) while recovering from infectious diseases, which could lead to permanent damage to hearing or vision (Lomax, 1996, pp. 118-9).

[31] Due to the imperfect diagnostic ability of physicians during the nineteenth century, some patients described as suffering from a disease of the joints (e.g. hip or knee) may in fact have had tuberculosis. We show that our results are not sensitive to the categorization or exclusion of admission for tuberculosis or diseases of joints.

attending physician. Doctors treated medically complicated and interesting cases – such as neurological disorders or severe infections – whereas surgeons and assistants treated patients with fractured bones or rheumatic conditions such as a diseased knee. Since the decision to assign a doctor to a patient was also guided by the severity of the condition, we consider treatment by a doctor to be a strong indicator of poor childhood health or a severe health shock.

Since information contained in the diversity of conditions recorded in the hospital registers is lost in the binary classifications, we also create a more continuous treatment variable to better exploit the variation in the causes of admission. For each unique condition or symptom recorded in the hospital registers, we compute the fraction of admissions that terminates with death in the hospital. Table A8 illustrates the construction of this continuous mortality index with an example of a cause of admission that contains three components: "Abd. pain, ?Enteric fever". This patient was admitted with abdominal pain that may have been caused by a case of typhoid (enteric) fever. We compute separate mortality rates for individuals admitted with enteric fever, any condition containing the words "abdominal" or "abdomen", and any condition with symptoms of pain.[32] In the main specification we assign an individual the highest mortality rate among all components in the cause of admission text string, but as a robustness check we also present estimates in Tables A10 and A11 from specifications using the lowest mortality rate among components or the mortality rate for the most frequently observed condition.

4 Estimation Framework

4.1 Empirical specifications

We estimate a sibling (household) fixed effects model to address the potential bias due to correlation between unobserved household characteristics, health status, and the socioe-

[32] See Appendix 9 for more details on how the mortality rates are computed.

conomic outcomes. The identifying assumption in the household fixed effects model is that there are no unobserved sibling-specific factors that are correlated with hospitalization and explain the outcomes of interest. In the simplest specification, the treatment is an indicator variable for admission to a hospital of sibling i from household j. We collapse the records of patients admitted to a hospital multiple times (11.5 percent of patients in our final sample) into a single observation so that each observation in the sample represents a unique individual. We examine the effects of childhood health deficiencies on occupational class, intergenerational occupational mobility, occupational log wages, and marriage market outcomes. In particular, we estimate the following equation:

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

where Y_{ij} is an occupational, intergenerational, or marriage market outcome for individual i from childhood household j, $Hospitalized_{ij}$ is an indicator for an individual who appears in the hospital admission registers, X_i is a vector of individual characteristics (inferred birth year and birth order fixed effects, in addition to an indicator for an outcome that is observed in 1911 as opposed to 1901), α_j denotes unobservable time-invariant determinants of the outcomes that are specific to a household, and ε_{ij} is a heteroskedasticity-robust error term clustered at the childhood household level that represents sibling-specific unobserved characteristics.[33] The coefficient of interest is β, which can be interpreted as the effect of a childhood health deficiency as proxied by hospitalization, relative to a sibling who does not appear in the hospital records.

This simple proxy for childhood health is arguably a crude measure as it lumps together admissions attributed to lower levels of health capital, as well as those that merely occurred because of accidents or other causes only weakly correlated with poor health. Therefore, in our preferred specification, we substitute the binary hospitalization treatment with the in-

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

hospital mortality index, a continuous measure of poor health computed from the detailed causes of admissions and described in Section 3.5. This index will mechanically bias the estimates towards zero, since children who survive long enough for us to observe their long-run outcomes will have above average levels of unobserved health capital. The remainder of the equation remains the same as Equation 1:

$$Y_{ij} = \beta \textit{In-hospital Mortality Rate}_{ij} + \gamma X_i + \alpha_j + \varepsilon_{ij} \tag{2}$$

As an alternative to the continuous measure of health deficiency we also use a hand-coded grouping of the conditions into "acute" or "non-acute" categories, which we described in detail in Section 3.5. We substitute the single treatment variable with two indicators, one that includes admissions related to poor childhood health and another that contains all other causes of admissions. As before, we always compare patients to their non-hospitalized siblings of the same gender. We estimate the following equation:

$$Y_{ij} = \beta_1 Acute_{ij} + \beta_2 Non\ acute_{ij} + \gamma X_i + \alpha_j + \varepsilon_{ij} \tag{3}$$

where *Acute* is an indicator variable for at least one admission with an acute condition, and *Non acute* is an indicator for any other hospital admission.[34] The coefficient of interest is β_1 which is interpreted as the effect of a health deficiency proxied by admissions for an acute condition, relative to a sibling who does not appear in the hospital records. We also expect β_2, the coefficient on non-acute conditions, to be zero. Finally, we separate admissions into those where the patient was treated by a doctor, and those treated by a surgeon or assistant medical provider. In particular, we estimate:

$$Y_{ij} = \beta_1 Doctor_{ij} + \beta_2 No\ doctor_{ij} + \gamma X_i + \alpha_j + \varepsilon_{ij} \tag{4}$$

[34]For 25.6 percent of patients admitted multiple times (or 2.9 percent of all patients), we observe at least one admission for an acute condition and at least one admission for a non-acute condition. In these cases, both variables take the value of one.

In this classification, cases treated by a doctor provide a proxy for poor childhood health that is not affected by the subjective classification of the causes of admission.

Our sample restricts attention to patients who resided in London and surrounding countries at the time of admission, in addition to non-hospitalized siblings of the same gender as a patient, who resided in the same household when enumerated in a census during childhood.[35] The coefficient β (or β_1 and β_2) is identified from households with one patient and one non-hospitalized sibling successfully matched to a census during adulthood. Our main specification requires that the average discrepancy in self-reported age across sources be less than or equal to 1 year, and, as described in Section 3.2, limits the number of records in the census with neighboring birth years and names similar to each matched observation. In Section 6.4, we present robustness checks and test the sensitivity of our results to these assumptions.

4.2 Identification and threats to the research design

Within the proposed sibling fixed effects framework we can identify the causal effect of poor childhood health on long-run socioeconomic outcomes under two assumptions. First, within-household comparisons account for all unobservable factors that could be correlated with poor health and explain the outcomes. Equations 1 to 4 essentially difference out any household-specific and time-invariant confounders and identify β based on between-sibling variation in health. Second, indicators for hospitalization, acute conditions, and treatment by a doctor, as well as the in-hospital mortality rate, are strong predictors of deficiencies in health. We can then interpret these effects as the cumulative deficiency in unobserved health capital, encompassing health conditions from the neo-natal period to the time of hospitalization.

The cumulative nature of this underlying variable arises from the fact that, unlike

[35] The counties of residence in the sample are London, Essex, Kent, Middlesex and Surrey, which account for 92.8 percent of all admissions and 94.3 percent of matched patients. Our results are unchanged when we include patients from all counties.

Currie *et al.* (2010), we do not observe anthropometric measures such as birth weight or gestational length. It may well be the case that patients become hospitalized due to poor neonatal and post-natal health. We do not interpret the binary treatment variable that restricts attention to acute conditions as a measure of severity since we cannot distinguish between a case of acute pneumonia attributed to a more virulent infection, on one hand, or weaker baseline health, on the other. However, the continuous treatment based on the in-hospital mortality rate can be interpreted as a measure of severity since it reflects differential mortality rates across a large set of conditions.

The underlying health variable also includes all adverse health shocks, such as exposure to infectious diseases, that patients experienced from birth to the time they were admitted to the hospital, as well as the net effects of hospital care. Thus, we cannot interpret β as the effect of an exogenously occurring illness immediately prior to hospitalization.[36] In other words, suppose that there is a certain health threshold for each child beyond which the child is sick enough to be admitted as a hospital inpatient. We assume that the patients cross this health threshold whereas their siblings, possessing sufficiently better underlying health, do not. Given that the sibling comparison accounts for all time-invariant characteristics of the household, and also partially for genetic and intergenerational health transmission, we attribute any observable differences to the differences in individual health capacity.

Although the sibling fixed effects strategy is appealing, there are still several sources of potential bias in our data. Given that we draw upon historical data and construct a panel dimension that links individuals at different stages of their life, ten to thirty years apart, measurement error is a major concern. In particular, selective mortality and out-migration from England could be problematic for our estimates.[37] Late-nineteenth century England

[36] As we described in Section 2.2, doctors had limited means to treat diseases beyond palliative care, and thus our estimates are not confounded by the benefits of hospital care to the same extent that they would with modern day hospitalization. Patients may still have benefitted relative to non-hospitalized siblings due to their removal from a crowded, polluted and disease-ridden home environment.

[37] Previous studies have produced mixed evidence on the relative strengths of the selective mortality and scarring effects of illness. Costa (2012) finds a scarring effect among Union Army prisoners of war who were imprisoned before age 30, but also shows that the effect of mortality selection dominates the scarring effect among those imprisoned after age 30, who faced a lower risk of older-age mortality.

was characterized by high infant mortality, with as many as 163 deaths per 1,000 live births in 1899, and hence by positive selection into childhood in terms of unobserved health capital. The in-hospital mortality rate for admissions up to age one exceeded 31 percent, before falling to 14 percent for admissions between ages two and three, 7 percent for admissions between ages four and seven, and 5 percent for admissions between ages eight and eleven.[38] If, for example, the health insults of patients hospitalized during infancy are severe enough to prevent them from surviving long enough to be enumerated in a census during adulthood, then we observe a positively selected sample of the strongest surviving patient, which would produce a smaller difference in comparison to unaffected brothers (Bozzoli *et al.*, 2009).[39]

Another source of measurement error is the fact that we do not observe the health status of the "control sibling" who does not appear in the hospital records. It is not possible to rule out that the "healthy" siblings also experienced childhood illness. Siblings of hospitalized patients faced a greater than average risk of illness given the likely spillover effects from disease transmission within the household. Parents may also have sent their children to a hospital for which the admission registers have not survived, or may only have been able to afford sending a single child to the hospital. These sources of measurement error would bias the estimated coefficient towards zero and a statistically significant negative coefficient

[38]Given the high mortality rate of infants, and support for a prevalent belief that infants should not be separated from the mothers, the official policy of the children's hospital (GOSH) was to turn away patients under the age of two years. However, the hospital admitted a small number of infants from its founding in 1852 before relaxing the admission requirements in the mid-1870s, after which point the share of admissions under the age of two increased to 25 percent at the end of the nineteenth century.

[39]However, positive selection among siblings would produce a bias consistent with our current results. Suppose, for example, that a set of siblings enumerated in the census during childhood consists of three brothers: one who ends up in the hospital due to poor health, one who has poor health, but is marginally healthier than the threshold beyond which the illness would force his hospitalization, and one who has very strong underlying health. If we compare the first two siblings, then our estimates are biased towards zero. However, if the middle sibling happens to die or is not matched between censuses in childhood and adulthood, then our comparison is between a sibling in poor health and a sibling with much better underlying health than the average health of non-hospitalized siblings in the family. This could potentially create a larger difference in outcomes than what one would expect from the "true effect." We deal with this potential bias in two ways. First, we only compare the patients to the matched sibling closest in age. Second, we provide a set of estimates in which we vary the upper bound of the spacing between patients and siblings in hopes that siblings 0 to 2 years apart in age from patients would unlikely to be subject to the described bias. The results presented in Figures A3 and A4 support the conclusion that, if this bias exists, it is rather small and should not alter our conclusions.

should be considered as a lower bound.

A separate source of bias could be introduced by parental responses to a child's health or other aspects of child development A large literature in economics describes the differential treatment of siblings by their parents, especially in developing countries. The sibling fixed effect would not capture these parental behaviors because they vary over time and are directed towards a particular child. Given that we compare same-gender siblings, any gender component of differential behavior would affect both siblings identically and would be captured by the fixed effect. Thus, we can abstract from gender discrimination within poor families. The parents in these household could essentially behave in three ways. First, they can be neutral and treat all children in the same way, which would not impose a bias in our estimates. Second, parents could compensate for adverse environmental or health shocks by directing resources away from the "stronger" child towards the relatively "weaker" child. In this case, the treated sibling would look more alike to his untreated brother, and we would underestimate the true effects of poor childhood health. Third, parents can reinforce the negative events by directing resources away from the relatively "weaker" child towards the "stronger" child. This behavior would be consistent with our negative estimates of poor childhood health as it would make the treated child appear worse than he would have been from lower health capital alone. Thus, reinforcement of health shocks by parents is a potential mechanism for the weaker occupational outcomes observed among children that experienced health insults.

5 Main Results

We now turn to the main regression results in which we estimate Equations 1 to 4, and compare hospitalized patients to their non-hospitalized siblings of the same gender. In Sections 5 to 7 we focus solely on males, while in Section 8 we analyze effects on female patients. Panel A of Table 4 presents estimates of sibling fixed effects regressions in which

the treatment of interest is an indicator variable for hospitalization (Equation 1), while panel B contains estimates in which the treatment is the continuous in-hospital mortality index (Equation 2). In Table 5, panel A shows estimates in which we include separate indicator variables for acute and non-acute conditions as diagnosed at admission (Equation 3), while panel B presents estimates from a specification with separate indicator variables for patients treated by a doctor and patients treated by a surgeon or assistant medical provider (Equation 4). In each case the comparison group is a non-hospitalized sibling.

Throughout the analysis we study six outcome variables, three of which reflect individual occupational success and can be thought of as a cumulative distribution function for occupational class: the probability of working in a white collar occupation (column (1)); the probability of working in a white collar or skilled occupation (column (2)); and the probability of working in a white collar, skilled, or low-skilled occupation (column (3)).[40] The next outcome (column (4)) is a continuous measure of occupational success in which we group occupations into one of twenty-two categories and assign each the annual occupational log wage based on Williamson (1980).[41] The final two outcome variables focus on intergenerational mobility in occupational attainment: the probability of attaining a lower occupational class than one's father (column (5)) and the probability of attaining a higher occupational class than one's father (column (6)). These intergenerational outcomes can be thought of as measures of occupational success relative to family endowments. The effects on a downgrade in occupational class are only identified from cases in which the patient attains a lower occupational class than his father, but the non-hospitalized sibling does not. Thus, the coefficient on hospitalization can be interpreted as the differential effect on the probability of downward mobility for patients relative to non-hospitalized siblings.

The results in panel A of Table 4 indicate that hospitalized children are 3.9 percentage

[40]The second dependent variable can also be interpreted as one minus the probability of working in a low-skilled or unskilled occupation, and the third dependent variable can also be interpreted as one minus the probability of working in an unskilled occupation.

[41]Each occupation is first assigned to one of five occupational classes based on Armstrong (1972). Occupational titles which do not fall under one of the twenty-two categories with a known occupational log wage are assigned the average occupational log wage for occupations in its occupational class.

points less likely to enter white collar occupation, 4.3 percentage points less likely to enter white collar or skilled occupation, and 13.4 percentage points more likely to enter unskilled occupations. These patients also earned occupational wages that were 4.2 percent lower than their brothers as adults. These coefficients imply effects on the upper half of the occupational distribution between 8 and 14 percent, and no statistically significant effect on the probability of entering unskilled occupations. When we focus on our preferred specification in panel B, in which we proxy for health insults using the continuous mortality index, the largest reductions in occupational success are also found for the probability of attaining white collar status. The coefficients decrease in magnitude when we add lower-class occupations to the dependent variable. Similarly, we find sizable and highly statistically significant negative effects of health insults on occupational log wages ten to thirty years later. In Table 6, we also present ordered probit estimates for the four occupational outcomes which align closely with our estimates from the linear probability model. Finally, in Table A5 we replicate Table 4 without sibling fixed effects. These estimates are generally smaller in magnitude which suggest that there are significant differences across household that correlate positively with both health status and long-run outcomes of children.

In Table 5, we present estimates from two alternative specifications for health deficiencies, in which we include separate indicator variables for admissions diagnosed as acute or non-acute conditions (panel A), as well as separate indicators for whether the condition was treated by the doctor or other medical professionals (panel B). In both the case of acute admissions and treatment by a doctor we find large negative effects on the probability of occupational success and on annual occupational log wages. For example, being admitted with an acute condition decreases the probability of working in a white collar occupation by 9 percentage points in comparison to one's brother, while treatment by a doctor decreases the probability by 7.3 percentage points. On the other hand, for conditions that should not be related to poor underlying health and should not yield long-run consequences, we do not find any large or statistically significant effects. These estimates for occupational attainment

range from -1.7 to 0.9 percentage points depending on the specification and the outcome variable. In fact, even with our relatively small sample sizes, we have enough power to reject the equality of the two treatments indicators (e.g. acute vs. non-acute). From these results we conclude that the negative long-run effects are driven by underlying health deficiencies rather than simply requiring admission to the hospital during childhood. The effects of acute conditions and treatment by a doctor are also non-negligible economically. For example, the decrease in probability of attaining white collar occupational status is between 27 and 33 percent, while the probability of working in either white collar or skilled occupation falls by about 18 percent. However, the percent effects for non-acute admissions or those treated by surgeons or medical assistants are very small and range from -5 to 3 percent.

Thus far, we have quantified the effects of health deficiencies on individual occupational success and occupational log wages in relation to the sample means. Although widely understood, this comparison does not yield a tangible economic interpretation. Therefore, in Table 7, we scale our effects by estimates of the intergenerational occupational elasticity, a measure often used in economics to understand social mobility and inequality across generations. In particular, we estimate the following equation:

$$(\textit{Son's status})_i = \gamma\,(\textit{Father's status})_i + \delta X_i + \epsilon_i \tag{5}$$

in which we regress the son's occupational status on the father's occupational status, and control for birth year fixed effects (for father and son) and birth order fixed effects (for the son).[42] In column (1), the dependent variable and regressor of interest are indicators for attaining white collar occupational status, for the father and son, respectively, and in column (2) they are indicators for reaching white collar or skilled occupational status. In column (3), we estimate an intergenerational wage elasticity by regressing the occupational log wage of the father on that of the son. Across the three specifications, we find elasticities

[42]We also control for the census year in which the son's outcome is observed. We restrict attention to fathers born between 1826 and 1865 who were between the ages of 26 and 55 when enumerated in the census.

ranging from 0.16 to 0.28. In panel B, we scale the estimates for the effects of health deficiencies from Tables 4 and 5 by the intergenerational elasticities. Most notably, we find that a one-standard-deviation increase in mortality index offsets 30.3 percent of the effect of a one-standard-deviation increase in father's occupational log wage.[43] The effects of acute conditions and treatment by a doctor, which isolate cases with significant health deficiencies, are much larger as they offset between 32 and 58 percent of the advantage of having a father with a higher occupational status.

In columns (5) and (6) of Table 4, we present our second set of main results which describe the effects of health deficiencies on intergenerational mobility. In particular, we estimate the effects on the probability of becoming more or less successful than one's father, with the treatment varying as shown in Equations 1 to 4.[44] We interpret the effect as a relative measure of own occupational success conditional on the occupational status of the father for a hospital patient in comparison to his brother (Chetty *et al.*, 2014). The results in panel A imply that children in poor health have a 2.2 percentage points greater probability than their brothers of achieving a lower occupational status relative to their father. Conversely, their chances of attaining a higher occupational status are 1 percentage point lower. Although these estimates are statistically insignificant their signs are consistent with our findings for probabilities of individual occupational success. They also appear to be asymmetric as we find the probability of downward mobility to be more than twice as large as the probability of upward mobility. In panel B we present our preferred estimates with the mortality index as a treatment variable that proxies for health deficiencies. In this case, we observe a similar pattern as in the binary treatment case, but the estimates are

[43] A one-standard-deviation change in the mortality index is 0.097, which has a 3.1 percent ($0.361 \times 0.097 \times 100$) effect on an individual's occupational log wage. Similarly, a one-standard-deviation-change in the father's occupational log wage is 0.283, which has an 11.5 percent ($0.283 \times 0.408 \times 100$) percent effect on the son's occupational log wage. Thus, health deficiencies offset 30.3 percent (3.1/11.5) of the effect of the father's occupational log wage.

[44] We exclude observations for which the father's occupation is missing, as well as cases where the father's occupation is in the highest or lowest occupational class (Classes 1 and 7 in the seven-class HISCLASS scheme). We exclude these cases because an individual with a father in a class 1 occupation (professional) would not be at risk of upward mobility.

statistically significant, implying that the more severe a condition for which an individual was admitted to the hospital, the higher the likelihood of downward mobility and the lower the chances of upward mobility – again, in comparison to one's brother and conditional on the father's occupational success. In panels A and B of Table 5, we also present the results for acute and non-acute admissions as well as admissions that were treated by a doctor as opposed to those treated by a surgeon or medical assistant. The results from these specifications support our main intergenerational findings and are consistent with the results for unconditional occupational success, namely that the effects are driven by conditions that we would expect to impose a burden on one's health, rather than by accidents or minor health issues. The estimates imply that health deficiencies are associated with a 22 to 23 percent increase in the probability of downward mobility and a 13 to 15 percent reduction in the probability of upward mobility.

To get a sense of how economically meaningful these estimates are, we can ask what fraction of overall intergenerational mobility can be explained by health deficiencies. In Table 8, we scale our main estimates by the rates of mobility in our sample and in the population. Panel A shows that the shares of our linked sample that experience downward and upward mobility are 0.30 and 0.35, respectively, while the corresponding figures for the population of England are 0.22 and 0.27. The estimates for the population are based on a linked sample from 1881 to 1901 (Long, 2013), and are much smaller than the raw mobility rates in our sample, which is restricted to the metropolitan environment of London, where the prospects of mobility were above average. Panel B reports the scaled health deficiency coefficients. In our preferred specification, the effect of a one-standard-deviation increase in the mortality rate accounts for 8 to 12 percent of the overall downward mobility, and 5 to 7 percent of observed upward mobility. As we found with the scaled estimates for the effects on unconditional occupational success, the binary treatments that restrict attention to admissions reflecting health deficiencies are much larger as they account for 22 to 32 percent of observed downward mobility.

6 Robustness Checks

6.1 Sample selection

In Table A9, we explore the extent to which our main results are affected by sample selection. In particular, we remove patients admitted for conditions with the 5 percent highest mortality rates (panel A), patients admitted with contagious diseases (panel B), patients admitted during infancy at ages 0 to 1 (panel C), and patients admitted multiple times (panel D). Our main results for the effects of health deficiencies, proxied by the mortality index, are qualitatively invariant to these changes in sample composition. In the case of individual occupational success, the estimates from these selected sub-samples are always greater in magnitude than the main estimates. They range from as little as 3 percent larger when we remove patients admitted as infants, to as much as 50 percent larger estimate when we remove contagious diseases, in the specification with the probability of attaining white collar status as the outcome. For effects on occupational log wages, we observe modest increases in magnitudes between 3 and 14 percent, but also a 17 percent decrease in the size of the estimate when we remove contagious diseases from the sample. Finally, for the intergenerational measures, we observe larger increases in the probability of downward mobility, ranging from 10 percent, when removing admissions during infancy, to 25 percent, when removing contagious diseases. The pattern is not so clear for the effects on the probability of upward mobility, for which the estimates are insignificant in three out of four cases. Overall, we find suggestive evidence that many of the biases related to selective mortality, or contamination of the control group through within-household transmission of health disadvantages, bias our main results downwards, and once these observations are removed, the estimates become larger in magnitude.

6.2 Specification of the mortality index and multiple admissions per patient

In our main specification, we construct the mortality index using data on all patients from the admission years and birth cohorts included in our main estimation sample. We also use the highest mortality rate when multiple conditions are included in the cause of admission. Since this particular choice of specification can be viewed as subjective, we test the sensitivity of our results to alternative methods of constructing the treatment variable. Table A10 investigates whether our results are affected by the choice of admissions used to calculate the mortality rates. Table A11 shows estimates from specifications that change the method of selecting the mortality rate when admissions report multiple conditions or symptoms. The results are qualitatively unchanged irrespectively of the sample that we use in constructing the mortality index, and, in quantitative terms, they change by no more than 25 percent. In fact, the magnitudes change by more than 10 percent in only four out of twelve cases. The results are also remarkably similar when we substitute the highest mortality across multiple conditions with either the lowest mortality rate (panel A), or the mortality rate of the most frequently reported condition (panel C). The estimate only become much larger when we use a mortality index weighted by the frequency of each condition (panel B). Although this specification produces magnitudes nearly double in size, it also yields substantially inflated standard errors in comparison to the main specification. Nonetheless, the conclusions remain qualitatively similar.

Our main specification is conservative since we collapse multiple admissions into a single observation per patient and ignore information on individuals with repeated admissions for the same condition. It is important to rule out whether the effects can be attributed to cases of children with multiple admissions for chronic diseases, which could reflect not only lower health capital, but also the direct negative effects of prolonged hospitalization and disability. Moreover, the set of patients admitted multiple times is likely to be highly selected, given that high-socioeconomic-status households were better able to afford repeated visits, and, in

some cases, doctors followed up with patients at home, eliminating the need for readmission.[45] In Table A9, we already tested that patients admitted multiple times are not driving our results. However, it is important for the interpretation of our findings to understand what happens to the estimates when we adopt different specifications of the treatment in cases when patients are admitted multiple times.

In Table A12, we restrict attention to the first observed admission (panel A) and we construct a count variable for the sum of all admissions with a given condition (panel B), which is directly comparable to the treatment in Currie *et al.* (2010). Our results in Table 4 change very little regardless of how we specify multiple admissions, and the magnitudes of the estimates are in the neighborhood of 30 percent for the white collar occupation, 17 percent when combining white collar and skilled occupations, 15 percent for downward mobility, and 8 percent for upward mobility, although the latter is insignificant. Thus, we conclude that our results do not reflect the outcomes for a selected group of patients admitted multiple times with chronic conditions, and are invariant to different specifications of the treatment variable.

6.3 Geographic selection

A potential concern with our estimates is that they may be biased by differential selection of patients based on the distance travelled to the hospital. The general hospital (Barts) and children's hospital (GOSH) in our sample admitted 16 and 20 percent of their patients, respectively, from outside the county of London. Providing care to children from outside of London was especially common at GOSH, which specialized in the treatment of congenital malformations and genetic disorders. Presumably, the mother of a patient faced a higher cost of bringing her child to the hospital the farther they travelled, and thus patients residing

[45] While high-SES mothers could turn to others in helping to raise their children, low-SES mothers were often the only providers of care for their children. Thus, the time cost of bringing a child to the hospital may have been higher for working class mothers. The children's hospital often discharged patients who contracted an infectious disease in the hospital and sent a Visiting Officer to care for the child at home. Some case notes also mention that doctors monitored the recovery of a patient by keeping in touch with the family after discharge (Tanner, 2007, pp. 146-158).

outside of London would potentially have larger health deficiencies.[46] Our main specification restricts the sample to patients from the Greater London area, but Table A13 shows estimates from a sample restricting attention to patients residing in the county of London (panel A), and from an expanded sample of all patients residing in England (panel B).[47] The results are very similar when we consider patients from anywhere in England, and they are generally larger when we focus on patients from the County of London, except for the effects on occupational log wages which are virtually identical. This finding is consistent with inner London having poorer living conditions than surrounding areas, as well as a positively selected sample of patients being admitted from the suburbs of London.

6.4 Match quality

The probabilistic matching procedures used to link various historical sources, such as the hospital registers and census records utlized in this paper, rely heavily on assumptions about the appropriate thresholds for considering an observation successfully matched. In our case, a crucial variable is the number of individuals enumerated in the census with names similar to, and with birth years within one year of, the census record matched to a hospital patient.[48] As we relax the threshold for the number of records similar to the matched observation, we trade off increased statistical power with a higher probability of a false positive match. In the absence of any formal tests to determine the appropriate thresholds, we present Figures A6a and A6b which show the main estimates with corresponding 95 percent confidence intervals plotted over samples that vary the degree of tolerance allowed for the accuracy of a match. In our main specification, we allow each individual in the census matched to a hospital admission to have up to ten other census records with similar names within one year of age. In the

[46] It is also the case that we are less likely to match individuals residing outside the County of London when admitted to the hospital.

[47] Recall that our definition of Greater London includes patients from the surrounding counties of Essex, Kent, Middlesex and Surrey.

[48] We consider the names of two individuals to be similar if the Levenshtein string distances of both the first name and surname, weighted by the length of each string, are less than 0.1. We also allow up to twenty census records with similar names within 2 years of age.

graphs, we vary the threshold for the number of similar records between zero and ten. For all outcomes of interest, the results are remarkably stable irrespectively of the restrictions placed on the matching criteria. The confidence intervals become moderately larger as the match criteria become more restrictive, which reflects the smaller sample sizes in these regressions. These results suggest that our main findings are not driven by measurement error related to false positive matches. If the false positive match rate increased as we allow more matches with non-unique names, then we should observed much different estimates based on the sample with restrictive matching criteria, in comparison to the sample with lenient criteria.

We present two additional robustness checks that address concerns about match quality in Table A14. In our main specification, we require individuals matched from childhood to adulthood census records to have reported the same birth county, but we allow the birth parish to be missing in one or more sources.[49] In panel A, we require the records linked across censuses to also match on birth parish, which potentially eliminates some false positives. In panel B, we restrict the sample to observations matched via "hints" to suggested records on the Ancestry.com genealogy website.[50] Our unconditional occupational success measures are generally larger in both cases and remain statistically significant, despite the fact that the sample sizes are reduced by about 50 percent. The results for the restricted sample of observations matched via the hints demonstrate the importance of measurement error, from matching records between censuses, that biases our results in the main specifications toward zero. In both panels of Table A14, the probability of downward mobility is smaller in comparison to the main estimates, and while these coefficients are no longer statistically significant, they suggest the same qualitative results. On the other hand, the effects on upward mobility are larger and in one case remains statistically significant.

[49] We don't require the birth parishes to match exactly across census because it is often missing in the 1911 census for individuals born in London, and geographic locations reported in 1911 did not always correspond to parishes that existed in earlier census years.

[50] In the main specifications, we link records between censuses either by entering search criteria for potential matches in the database of 1901 and 1911 census records, or by collecting links to the suggested records. As we described in section 3.2, we have reason to believe the observations matched via "hints" are more accurate than direct searches since some of these links are based on privately held information that does not appear in the census records.

A final robustness check, shown in Figures A3 and A4, examines whether our main estimates change as we vary the maximum age gap between patients and their siblings. This exercise addresses concerns that we do not match the closest sibling, due to selective morality, or that siblings born further apart face different environments within the household. The former would generate a downward bias, but the latter would produce bias of ambiguous sign as it violates the fixed effects assumption that all within-household components are captured by fixed effects. On the other hand, a sibling close in age to the patient could impose a larger burden on household resources, especially while the patient is sick, which limits the external validity of these comparisons. Figures A3 and A4 provide estimates for the effects of admissions with acute condition, similar to panel A in Table 4, and estimates for the effects of the in-hospital mortality index, akin to panel B in Table 5, respectively. In the main estimation sample, we include siblings to up to eight years apart in age from the patients, while in the graphs we vary the threshold between two and eight years. Although the confidence intervals are larger when we restrict attention to siblings close in age to the patients, our main estimates hold for all outcomes, and in fact, they are very similar across the range of age thresholds. This finding suggests that selective mortality or time-varying household shocks do not alter our results substantively.

We believe that the preceding discussion has provided compelling evidence that our main findings showing the negative effects of poor childhood health on individual occupational success and intergenerational mobility are qualitatively robust to many different specifications. In virtually all cases, we estimate significant negative effects of health deficiencies on the probability of entering white collar or skilled occupations. In fact, we often estimate negative effects that are larger in magnitude, in samples which exclude observations that may bias our main estimates towards zero and threaten identification, as discussed in Section 4.2. Given that our results appear robust to many reasonable specification checks, we move on to present heterogeneity analysis for our main specification in the following section.

7 Heterogeneity

Historical studies based on census data are typically unable to perform detailed heterogeneity analyses due to the limited number of auxiliary variables available, and our study is not an exception to this rule. Nonetheless, we present a few basic sample splits that provide some suggestive evidence of how our results vary across sub-populations. First, in Table A15, we present results separately for the two hospitals in our sample. The selection of patients admitted at each hospital may differ, given that Great Ormond Street Hospital was dedicated solely to child care, while St. Bartholomew's was a general hospital that admitted children. Second, in Table A16, we estimates separate effects for patients admitted during early childhood (ages 0 to 4), and those admitted during early adolescence (5 to 11). This exercise addresses not only the literature on early-childhood interventions and health shocks, but also the fact that the estimates for the younger patients may be biased downwards due to selective mortality.[51] Third, we investigate whether our results vary by the length of stay in the hospital, which may proxy for the severity of the health deficiency. Finally, we also investigate whether there is heterogeneity in the treatment effects by estimating separate indicators for each tercile of the mortality index.

During the late-nineteenth century, children stayed in hospitals for a much longer duration than is typically the case today, with median and mean lengths of stay of nineteen and thirty days, respectively. It is, however, an empirical question whether a longer stay indicated a more severe and potentially chronic disease, on the one hand, or provided benefits from longer exposure to improved nutrition, better sanitation, and a lack of overcrowding in comparison to the home environment, on the other. In Figure 8, we present estimates split by the terciles for the length of stay in our data. We find similar effects for the probability of attaining white collar occupation irrespectively of the length of stay, but in the case of occupational

[51] In fact, as documented in Table A9, the estimates become larger in magnitude when we exclude hospitalizations during infancy (ages 0 to 1), a change that is likely driven by a greater degree of selective mortality during infancy. If this bias is severe we would not be able to distinguish between selective mortality and weak effects during early childhood as reasons for finding smaller effects for younger patients, as opposed to those admitted later in childhood.

log wages and downward mobility, the results are clearly driven by the prolonged stays. However, we find an inverse u-shaped relationship for the effects on upward mobility, with the largest estimates occurring in the case of relatively shorter hospital stays.

In Table A16, we present a specification in which we estimate separate effects for patients admitted before the age of five, and those admitted on or after their fifth birthday, with results for the mortality index treatment in panel A and acute admissions in panel B. Due to the power constraints related to small sample sizes, we interact the treatment with age-group dummies instead of splitting the sample into two groups. Contrary to evidence on the importance of early-childhood health shocks, we find smaller and statistically insignificant results for the early-childhood admissions, in comparison to the hospitalizations later in childhood, in the case of the mortality index treatment. This pattern holds for acute admissions, with the exception of the downward mobility outcome, in which case we estimate a larger and marginally significant effect for the sample with patients admitted between ages zero and four. As noted in the introductory paragraph to Section 7, these results could reflect the importance for health capacity of the late-childhood period, relative to the early-childhood period, or perhaps more likely, indicate that the estimates for early-childhood admissions are biased towards zero due to selective mortality. Overall, these heterogeneity results suggest that, in addition to health conditions during early-childhood, health deficiencies during adolescence can also be detrimental to long-run occupational outcomes.

Thus far, we have pooled together patients from Barts and GOSH due to power constraints. However, in Table A15, we interact indicators for admission to a particular hospital with the indicator for acute conditions, in panel A, and the mortality index, in panel B. In both cases, we find that the results are driven mostly by children admitted at Barts as opposed to GOSH. The one exception to that pattern is the effect on occupational log wages, in which case we find comparable estimates for the two hospital samples.

In the last heterogeneity exercise, we estimate separate effects by tercile of the mortality index, our preferred treatment variable, to proxy for the magnitude of the health disadvan-

tage faced by the child. By construction, the mortality index biases us towards zero, and thus, we would expect the largest bias to be among the highest mortality conditions. The results presented in Figure 7 indicate that the effects of individual occupational success and intergenerational mobility are largest among the most severe health insults. The conditions in the top tercile of the mortality index included the major infectious diseases and their sequela, conditions described as acute, and injuries such as burns or fractured skulls. These results suggest that more severe health deficiencies are associated with a worse outlook in the long run.

8 Effects on female patients

Previous work based on linked historical census data is almost exclusively limited to samples containing only males.[52] An innovation of this paper is to link the childhood census records of females to a database of Church of England marriage licenses, which crucially contain a bride's maiden name and her spouse's surname, as well as both fathers' names. The additional information in the marriage records makes it possible to link the childhood records of married females to census records during adulthood. Since we have only digitized the records of female patient admissions to Great Ormond Street Hospital, we necessarily restrict our analysis to the sample of patients from the children's hospital.

8.1 Marriage markets in Victorian England

By the middle of the nineteenth century, England and Wales had settled into a relatively stable marriage pattern with a mean age at first marriage of twenty-five years, while 10 percent of men and 14 percent of women never married (Woods, 2000, p. 81).[53] While

[52]The only paper of which we are aware that uses linked historical data and includes female observations is Parman (2010) who links North Carolina death certificates of males and females to childhood census records using the mother's maiden name reported in the death certificates.

[53]It was common for spinsters to have a career, but married women usually gave up employment (Savage and Miles, 1994, p. 31).

Londoners typically married two to three years younger than the national average, the mean for the metropolitan area obscures significant contrasts in the marriage patterns between wealthy and poor districts. The wealthy married later and were more likely to remain single than the working-class.[54] Occupational choice accounted for a significant portion of the variation in the marriage pattern of males, as the mean age at first marriage ranged from 24.1 years for miners (with a mean age for spouses of 22.5 years) to 31.2 years for professionals (26.4 years for spouses) in the mid-1880s (Woods, 2000, p. 86).

Although family members generally had less influence on the choice of spouse than the choice of career by children, girls faced greater scrutiny from parents during the period of courtship. The necessity of a child's wage to the household budget could also force courting couples to delay the decision to marry by many years (Ross, 1993, p. 67). Since a daughter's marriage to a husband with property or capital could provide insurance against financial hardship, working-class parents would investigate the financial prospects of suitors (Miles, 1999, p. 161). Prospective brides would have been concerned about the insecure earnings power of working-class men in London, the majority of whom were employed in the undercapitalized secondary sector, often in seasonal positions, earning wages insufficient to provide support in old age.[55] Moreover, illness or injury to the primary breadwinner could leave a household dependent on the wages of wives and children to avoid falling into chronic poverty, a state in which 30 percent of London's population was trapped in the late-1880s (Ross, 1982, p. 576-7).

8.2 Marriage records

A database of marriage licenses from Church of England parish registers in the London area provides a wealth of information on marriage market outcomes (Ancestry.com, 2010).

[54] In the 1881 census, only 14 percent of women in upper-middle-class Hampstead had married by the age of twenty-four, compared to nearly 50 percent in working-class Bethnal Green. By the ages of forty-five to fifty-four, 30 percent of women in Hampstead remained unmarried, in contrast to only 7 percent in Bethnal Green (Ross, 1993, p. 59).

[55] See Stedman Jones (1971) for a classic account of the "casual labor" problem – the over-population of extremely poor people – faced by London in the first half of the nineteenth century.

Each license contains a bride's maiden name, her age, the groom's name and age, both fathers' names, and the parish in which the marriage took place.[56] The most basic marriage market outcome is an indicator for marital status. If childhood illness has a long-lasting effect on earnings ability and attractiveness in the marriage market, then it is expected that the hospital patients will be less likely to marry than their siblings. For females, marriage in Victorian England can be viewed as a proxy for socioeconomic status as it provided women with a safety net. Given that the vast majority of married women did not participate in the labor market, we do not explore effects on their occupational and intergenerational outcomes.[57]

Several papers have used the occupational titles reported in marriage registry data to provide evidence of differential mobility by gender in nineteenth-century Britain. In particular, the consensus view of the British historical literature holds that a daughter's mobility in the marriage market, measured by her spouse's occupational class relative to her father's, was greater than the mobility of a son relative to his father in the labor market (Miles, 1999; Mitch, 1993). Empirical analysis based on marriage records alone has a number of limitations, the most prominent of which is that it excludes individuals who never married and ignores occupational upgrading that occurred after marriage.[58] Our paper addresses these issues by combining the marriage registry data with linked census records.

A related literature has studied the determinants of marriage market outcomes, such as the degree of assortative matching (Weiss, 1997). Abramitzky *et al.* (2011) exploit heterogeneity in World War I military mortality among soldiers born in different regions of France to show that a decrease in the population sex ratio (of men to women) reduced the probabil-

[56]The age of the bride and groom were recorded consistently starting in the mid-1870s, before which the licenses only reported that an individual had reached "full age." The licenses also report the occupational titles of the four individuals named in the record, but unfortunately these fields have not been transcribed.

[57]In our data, we can also investigate effects on the occupational attainment of the spouses of married females, as well as childbearing for both married and single females. However, our limited sample size prevents us from restricting attention to married females.

[58]Moreover, the records only include marriages that took place in the Anglican church, resulting in an unrepresentative sample of marriages after the turn-of-the-twentieth century. By 1914, civil ceremonies, Catholics, and Nonconformists (Lutherans, Jews, Quakers, etc.) accounted for 40 percent of marriages (Savage and Miles, 1994, p. 31).

ity that a groom would marry a bride from a lower class. In a related paper on the London Season, the exclusive marriage market of the Victorian-era British elite, Goni (2013) finds that the cohort of women affected by an exogenous interruption to the marriage market was 80 percent more likely to marry a commoner.

8.3 Data linkage

Figure 9 illustrates the linking procedure for the female sample. It essentially replicates the strategy from the analysis of male patients, depicted in Figure 3 and described in Section 3.2, with the exception of an additional linkage required to match females who changed their surnames at marriage. The childhood census records of female patients and their siblings are linked to a database of London marriage licenses by entering the bride's full name and her father's given name as search criteria.[59] Next, the marriage records are linked to the 1901 or 1911 censuses of England, at which point the individuals would have been adults, with a woman's first name and her spouse's full name entering as search criteria.[60] Finally, the linking procedure is completed by restricting attention to observations with childhood and adulthood census records that have matching birthplaces and deviations in years of birth no more than three years.[61]

The linked data set of census and marriage records provides two significant advantages over the analysis of social mobility by Miles (1993, 1999) and Mitch (1993) based on the marriage records alone. First, the linked data contain both married and single individuals, which permits an analysis of the propensity to marry. Second, the occupations of daughters and son-in-laws in the census are reported at older ages, which may provide a more accurate

[59] We limit searches to the London-area marriage licenses, given the time cost of performing the searches and the low probability of successful matches to marriage records from the rest of England. However, we obtain some observations with marriages that took place outside of London from links ("hints") to marriage records that appear on the pages of the childhood census records.

[60] We also verify that successful matches do not report conflicting values for years of marriage in the marriage license data and the 1911 census.

[61] An additional year of tolerance is allowed in the reported year of birth, compared to the linkage of the hospital records to the childhood census, due to the larger gap in time between the two sources and the greater probability of age heaping at older ages. See Baten *et al.* (2013) for a discussion of age heaping in the censuses of England.

measure of socioeconomic status than the occupational titles reported at the time of marriage, when individuals may not have reached the peak of their professional careeer. One remaining limitation of our linked data set of females is that the sample of married females is restricted to marriages that occurred in the Church of England.[62] We partially address this data limitation by including observation matched to adult census records using the "hints" on the genealogical website, which may contain civil and Nonconformist marriages. The issue of religious affiliation is also less of a concern for our study since London had one of the lowest proportions of civil marriages in England and Wales at 10 percent in 1884, before rising to 14 percent in 1894, 21 percent in 1904, and 27 percent in 1913 (Anderson, 1975).

8.4 Results on marital status of females

Table 9 presents the results for the effects of poor childhood health on marital status as adults for female patients and their sisters, as well as for male patients and their brothers. Panel A reports the effects of admission to the hospital, and panel B displays the effects of acute and non-acute conditions. For both definitions of the treatment, we find small and statistically insignificant effects of poor health on the probability of being married for males, but we find large penalties for females. These estimates are economically meaningful as they imply that female patients faced a 26 to 31 percent marriage penalty relative to their sisters.

9 Conclusion

This paper has documented the long-run scarring effects of poor childhood health using newly collected data on hospital admissions in late-nineteenth century London, England. Similar to Currie *et al.* (2010), we use hospitalization during childhood as a proxy for underlying health deficiencies and estimate its effects on occupational outcomes for males, and marriage market outcomes for both males and females. We present three main empirical

[62] The male sample includes individuals of all religious affiliations since the matching does not involve the marriage licenses.

findings.

First, a one standard deviation increase in the mortality index of the admitted condition is associated with a 0.068 standard- deviation decrease in occupational log wages for male patients as adults in comparison to their brothers. In particular, the patients were less likely to attain white collar or skilled occupational status. Second, the aforementioned increase in the mortality index is associated with a 0.054 standard deviation increase in the probability of downward mobility relative to one's father, for patients in comparison to their brothers. Moreover, the effect of health deficiencies proxied by the mortality index offsets 30.3 percent of the effect of a one-standard-deviation increase in the father's occupational log wage on own occupational log wage as adults. Third, we find that female patients were 8.6 percentage points (26 percent) less likely to be married as adults in comparison to their sisters, but we find no evidence of a marriage penalty for males.

These findings contribute to understanding the health deficiencies faced by children in late-Victorian England, and suggest that the consequences of poor health in a pre-modern disease environment encompass not only high rates of infant mortality, but also significant and economically meaningful effects on the long-run socioeconomic opportunity of survivors. While the decline in infant mortality in England during the early-twentieth century would typically be thought of as having positive effects on society, it also meant that the weakest children would now survive, when in the past they may have died during infancy. Our results suggest that for a significant segment of the population, the interaction of poor birth endowment, poor living conditions, and poor health in childhood made it highly probable that they would experience a lower quality of life in adulthood, and potentially require greater support from family and the state in old age. Our results on intergenerational mobility also suggest that poor childhood health promoted downward social mobility and prevented opportunities to advance up the occupational ladder. This finding provides quantitative evidence that health deficiencies in childhood can explain in part the intergenerational persistence of poverty.

Our results also have important policy implications for developing countries with high rates of infectious disease and infant mortality. In particular, access to health care facilities and medical practitioners could have substantially larger benefits for under-served, low-SES groups than previously thought, to the extent that it mitigates the effects of poor childhood health on long-run economic outcomes.. Moreover, focusing public health interventions on the weakest children, who in the past would likely have died during infancy, may have the greatest benefits for society. The findings also present a challenge to policymakers, as low levels of individual productivity at present can be explained in part by irreversible insults to health that occurred far into an individual's past.

An important question raised by our results is whether the long-run consequences of poor childhood health for the occupational outcomes of individuals translated into weaker macro-level economic performance for London and England. Although Acemoglu and Johnson (2007) find no causal evidence that large improvements in life expectancy increased income per capita, in a study of the global diffusion of drug and chemical innovations during the 1940s, they rely on gains to life expectancy at the end of life, whereas more substantial benefits to economic growth may be found with improvements in life expectancy that can be attributed to reductions in infant and child mortality. Conversely, the infant mortality rates that remained high until the end of the nineteenth century, and the debilitating childhood diseases that could limit educational attainment and cognitive ability, may have imposed long-run economic costs on England.

References

Abel-Smith, B. (1964). *The Hospitals, 1800-1948: A Study in Social Administration in England and Wales*. Cambridge, MA: Harvard University Press.

Abramitzky, R., Boustan, L. P. and Eriksson, K. (2012). Europe's tired, poor, huddled masses: Self-selection and economic outcomes in the Age of Mass Migration. *American Economic Review*, **102** (5), 1832–56.

—, Delavande, A. and Vasconcelos, L. (2011). Marrying up: The role of sex ratio in assortative matching. *American Economic Journal: Applied Economics*, **3** (3), 124–57.

Acemoglu, D. and Johnson, S. (2007). Disease and development: The effect of life expectancy on economic growth. *Journal of Political Economy*, **115** (6), 925–985.

Aizer, A., Eli, S., Ferrie, J. and Lleras-Muney, A. (2013). The effects of childhood means-tested cash transfers on mortality. *Working Paper*.

Almond, D. (2006). Is the 1918 influenza pandemic over? Long-term effects of in utero influenza exposure in the post-1940 U.S. population. *Journal of Political Economy*, **114** (4), 672–712.

— and Currie, J. (2011a). Human capital development before age five. *Handbook of Labor Economics*, **Volume 4, Part B**, 1315–1486.

— and — (2011b). Killing me softly: The fetal origins hypothesis. *Journal of Economic Perspectives*, **25** (3), 153–72.

Ancestry.com (2005a). *1891 England Census [Database on-line]*. Provo, UT: Ancestry.com Operations Inc.

— (2005b). *1901 England Census [Database on-line]*. Provo, UT: Ancestry.com Operations Inc.

— (2010). *London, England, Marriages and Banns, 1754-1921 [Database on-line]*. Provo, UT: Ancestry.com Operations, Inc.

— (2011). *1911 England Census [Database on-line]*. Provo, UT: Ancestry.com Operations, Inc.

Anderson, O. (1975). The incidence of civil marriage in Victorian England and Wales. *Past & Present*, (69), 50–87.

Armstrong, W. (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, England: Cambridge University Press.

Baird, S., Hicks, J. H., Kremer, M. and Miguel, E. (2012). Worms at work: Long-run impacts of child health gains. *Working Paper*.

Ball, M. and Sunderland, D. (2001). *An Economic History of London, 1800-1914*. New York: Routledge.

Barker, D. J. P. (1990). The fetal and infant origins of adult disease. *BMJ*, **301** (6761), 1111–1111.

Baten, J., Crayen, D. and Voth, H.-J. (2013). Numeracy and the impact of high food prices in industrializing Britain, 1780-1850. *Review of Economics and Statistics*.

Bhalotra, S. and Venkataramani, A. S. (2012). Shadows of the captain of the men of death: Early life health interventions, human capital investments, and institutions. *Working Paper*.

Black, S. E., Devereux, P. J. and Salvanes, K. G. (2007b). From the cradle to the labor market? The effect of birth weight on adult outcomes. *The Quarterly Journal of Economics*, **122** (1), 409–439.

Bleakley, H. (2007). Disease and development: Evidence from hookworm eradication in the American South. *The Quarterly Journal of Economics*, **122** (1), 73–117.

— (2010). Malaria eradication in the Americas: A retrospective analysis of childhood exposure. *American Economic Journal: Applied Economics*, **2** (2), 1–45.

Bozzoli, C., Deaton, A. and Quintana-Domeque, C. (2009). Adult height and childhood disease. *Demography*, **46** (4), 647–669.

Brown, R. and Thomas, D. (2013). On the long term effects of the 1918 U.S. infuenza pandemic. *Working Paper*.

Carpenter, M. W. (2010). *Health, medicine, and society in Victorian England*. Santa Barbara, Calif.: Praeger.

Case, A. and Paxson, C. (2008). Stature and status: Height, ability, and labor market outcomes. *Journal of Political Economy*, **116** (3), 499–532.

— and — (2010). Causes and consequences of early-life health. *Demography*, **47** (1), S65–S85.

Chetty, R., Hendren, N., Kline, P. and Saez, E. (2014). Where is the land of opportunity? the geography of intergenerational mobility in the United States. *National Bureau of Economic Research Working Paper Series*, **No. 19843**.

Condrau, F. and Worboys, M. (2007). Second opinions: Epidemics and infections in nineteenth-century Britain. *Social History of Medicine*, **20** (1), 147–158.

— and — (2009). Epidemics and infections in nineteenth-century Britain. *Social History of Medicine*, **22** (1), 165–171.

Costa, D. L. (2012). Scarring and mortality selection among Civil War POWs: A long-term mortality, morbidity, and socioeconomic follow-up. *Demography*, **49** (4), 1185–1206.

Currie, J., Stabile, M., Manivong, P. and Roos, L. L. (2010). Child health and young adult outcomes. *Journal of Human Resources*, **45** (3), 517–548.

— and Vogl, T. (2013). Early-life health and adult circumstance in developing countries. *Annual Review of Economics*, **5** (1), 1–36.

Farlex (2014). Thefreedictionary [online].

Ferrie, J. P., Rolf, K. and Troesken, W. (2012). Cognitive disparities, lead plumbing, and water chemistry: Prior exposure to water-borne lead and intelligence test scores among World War Two U.S. Army enlistees. *Economics & Human Biology*, **10** (1), 98–111.

Figlio, D., Guryan, J., Karbownik, K. and Roth, J. (2014). The effects of poor neonatal health on children's cognitive development. *American Economic Review*, **104** (12), 3921–55.

Floud, R., Fogel, R. W., Harris, B. and Hong, S. C. (2011). *The Changing Body : Health, Nutrition, and Human Development in the Western World since 1700*. Cambridge: Cambridge University Press.

Fogel, R. W. (2004). *The Escape from Hunger and Premature Death, 1700-2100*. Cambridge: Cambridge University Press.

Franklin, A. W. (1964). Children's hospitals. In F. N. L. Poynter (ed.), *The evolution of hospitals in Britain*, London: Pitman, pp. 103–121.

Gertler, P. and Gruber, J. (2002). Insuring consumption against illness. *American Economic Review*, **92** (1), 51–70.

Gluckman, P. D. and Hanson, M. A. (2006). The developmental origins of health and disase: an overview. In P. D. Gluckman and M. A. Hanson (eds.), *Developmental origins of health and disease*, Cambridge: Cambridge University Press.

Heckman, J. J. (2007). The economics, technology, and neuroscience of human capability formation. *Proceedings of the National Academy of Sciences*, **104** (33), 13250–13255.

Higgins, T. T. (1952). *'Great Ormond Street', 1852-1952*.

Hong, S. C. (2007). The burden of early exposure to malaria in the United States, 1850-1860: Malnutrition and immune disorders. *The Journal of Economic History*, **67** (04), 1001–1035.

— (2011). Malaria: An early indicator of later disease and work level. *Working Paper*.

Horrell, S., Humphries, J. and Voth, H.-J. (2001). Destined for deprivation: Human capital formation and intergenerational poverty in nineteenth-century England. *Explorations in Economic History*, **38** (3), 339–365.

Jayachandran, S., Lleras-Muney, A. and Smith, K. V. (2010). Modern medicine and the twentieth century decline in mortality: Evidence on the impact of sulfa drugs. *American Economic Journal: Applied Economics*, **2** (2), 118–46.

Kelly, E. (2011). The scourge of Asian flu. *Journal of Human Resources*, **46** (4), 669–694.

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

Leeuwen, M. H. D. v. and Maas, I. (2011). *Hisclass: A Historical International Social Class Scheme*. Leuven, Belgium: Leuven University Press.

—, — and Miles, A. (2002). *Hisco: Historical International Standard Classification of Occupations*. Leuven, Belgium: Leuven University Press.

Levene, A. (2012). *The childhood of the poor: welfare in eighteenth-century London*. New York: Palgrave Macmillan.

Lin, M.-J. and Liu, E. M. (2014). Does in utero exposure to illness matter? The 1918 influenza epidemic in Taiwan as a natural experiment. *National Bureau of Economic Research Working Paper Series*, **No. 20166**.

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

Long, J. (2013). The surprising social mobility of Victorian Britain. *European Review of Economic History*, **17** (1), 1–23.

— and Ferrie, J. (2013). Intergenerational occupational mobility in Great Britain and the United States since 1850. *American Economic Review*, **103** (4), 1109–37.

Medvei, V. C. and Thornton, J. L. (1974). *The Royal Hospital of Saint Bartholomew, 1123-1973*. London: Libraries Saint Bartholomew's Hospital Medical College.

Miguel, E. and Kremer, M. (2004). Worms: Identifying impacts on education and health in the presence of treatment externalities. *Econometrica*, **72** (1), 159–217.

Miles, A. (1993). How open was nineteenth-century British society? Social mobility and equality of opportunity, 1839-1914. In A. Miles and D. Vincent (eds.), *Building European Society: Occupational Change and Social Mobility in Europe, 1840-1940*, New York, NY: Manchester Univerity Press, pp. 18–39.

— (1999). *Social Mobility in Nineteenth- and Early Twentieth-Century England*. Houndmills, Basingstoke, Hampshire: Macmillan.

Mill, R. and Stein, L. C. (2012). Race, skin color, and economic outcomes in early twentieth-century America. *Working Paper*.

Minnesota Population Center (2008). North Atlantic Population Project: Complete count microdata version 2.0 [machine-readable database].

Mitch, D. (1993). 'Inequalities which every one may remove': Occupational recruitment, endogamy, and the homogeneity of social origins in Victorian England. In A. Miles and D. Vincent (eds.), *Building European Society: Occupational Change and Social Mobility in Europe, 1840-1940*, New York, NY: Manchester University Press, pp. 140–164.

Mokyr, J. (2000). Why "more work for mother?" Knowledge and household behavior, 1870-1945. *The Journal of Economic History*, **60** (1), 1–41.

Mooney, G. (2007). Infectious diseases and epidemiologic transition in Victorian Britain? Definitely. *Social History of Medicine*, **20** (3), 595–606.

Olivetti, C. and Paserman, M. D. (2013). In the name of the son (and the daughter): Intergenerational mobility in the United States, 1850-1930. *Working Paper*.

Oreopoulos, P., Stabile, M., Walld, R. and Roos, L. L. (2008). Short-, medium-, and long-term consequences of poor infant health. *Journal of Human Resources*, **43** (1), 88–138.

Parman, J. (2010). Gender and intergenerational mobility: Using health outcomes to compare intergenerational mobility across gender and over time. *Working Paper*.

— (2013). Childhood health and human capital: New evidence from genetic brothers in arms. *Working Paper*.

Ross, E. (1982). "Fierce questions and taunts": Married life in working-class London, 1870-1914. *Feminist Studies*, **8** (3), 575–602.

— (1993). *Love and Toil: Motherhood in Outcast London, 1870-1918*. New York: Oxford University Press.

Royer, H. (2009). Separated at girth: U.S. twin estimates of the effects of birth weight. *American Economic Journal: Applied Economics*, **1** (1), 49–85.

Savage, M. and Miles, A. (1994). *The Remaking of the British Working Class, 1840-1940*. New York: Routledge.

Schurer, K. and Woollard, M. (2003). National sample from the 1881 Census of Great Britain [computer file].

Seidler, E. (1989). An historical survey of children's hospitals. In L. Granshaw and R. Porter (eds.), *The Hospital in History*, New York: Routledge, pp. 181–195.

Smith, J. P. (2009). The impact of childhood health on adult labor market outcomes. *Review of Economics and Statistics*, **91** (3), 478–489.

Stedman Jones, G. (1971). *Outcast London: A Study in the Relationship between Classes in Victorian Society*. Oxford: Clarendon Press.

Strange, J.-M. (2005). *Death, Grief and Poverty in Britain, 1870-1914*. New York: Cambridge University Press.

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

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

— and Hawkins, S. (2013). *Myth, Marketing, and Medicine: Life in British Children's Hospitals , 1850-1914*, Bern, Switzerland: Peter Lang, pp. 209–236.

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

Venkataramani, A. S. (2012). Early life exposure to malaria and cognition in adulthood: Evidence from mexico. *Journal of Health Economics*, **31** (5), 767–780.

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

Weiss, Y. (1997). The formation and dissolution of families: Why marry? Who marries whom? And what happens upon divorce. In R. R. Mark and S. Oded (eds.), *Handbook of Population and Family Economics*, vol. Volume 1, Part A, Elsevier, pp. 81–123.

Williamson, J. G. (1980). Earnings inequality in nineteenth-century Britain. *The Journal of Economic History*, **40** (3), 457–475.

Wohl, A. S. (1977). *The Eternal Slum: Housing and Social Policy in Victorian London*. Montreal: McGill-Queen's University Press.

— (1983). *Endangered Lives: Public Health in Victorian Britain*. Cambridge, MA: Harvard University Press.

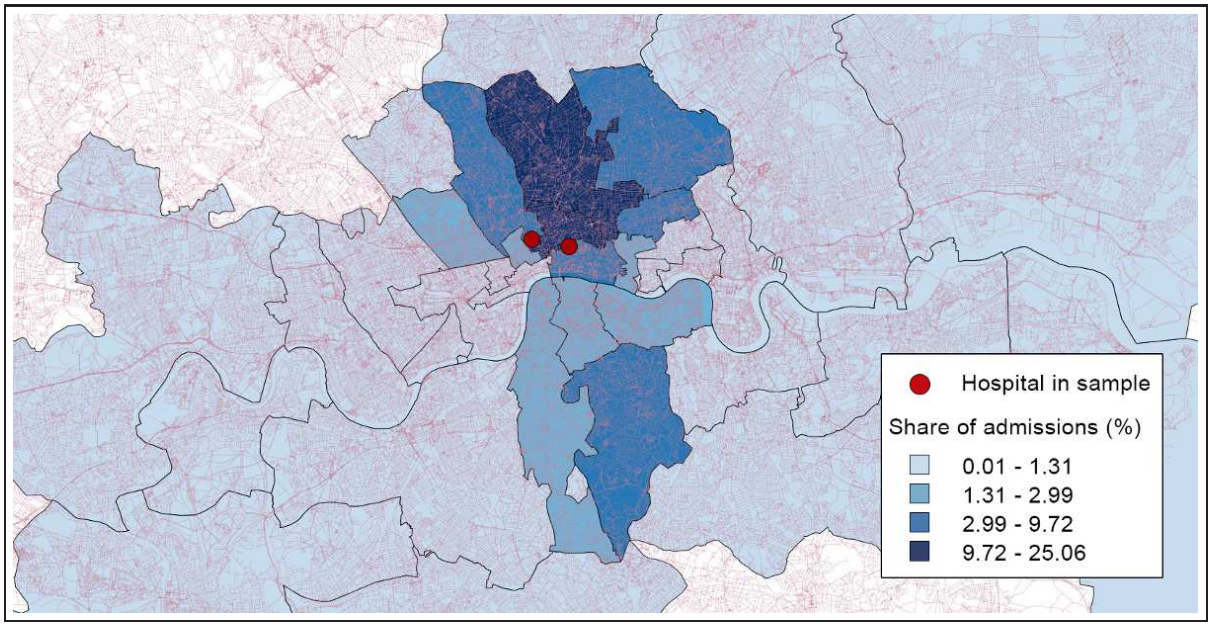
Woods, R. (2000). *The Demography of Victorian England and Wales*. Cambridge: Cambridge University Press.

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

Zhang, X. (2014). Children of the mortality revolution: Infectious disease and long-run outcomes. *Working Paper*.

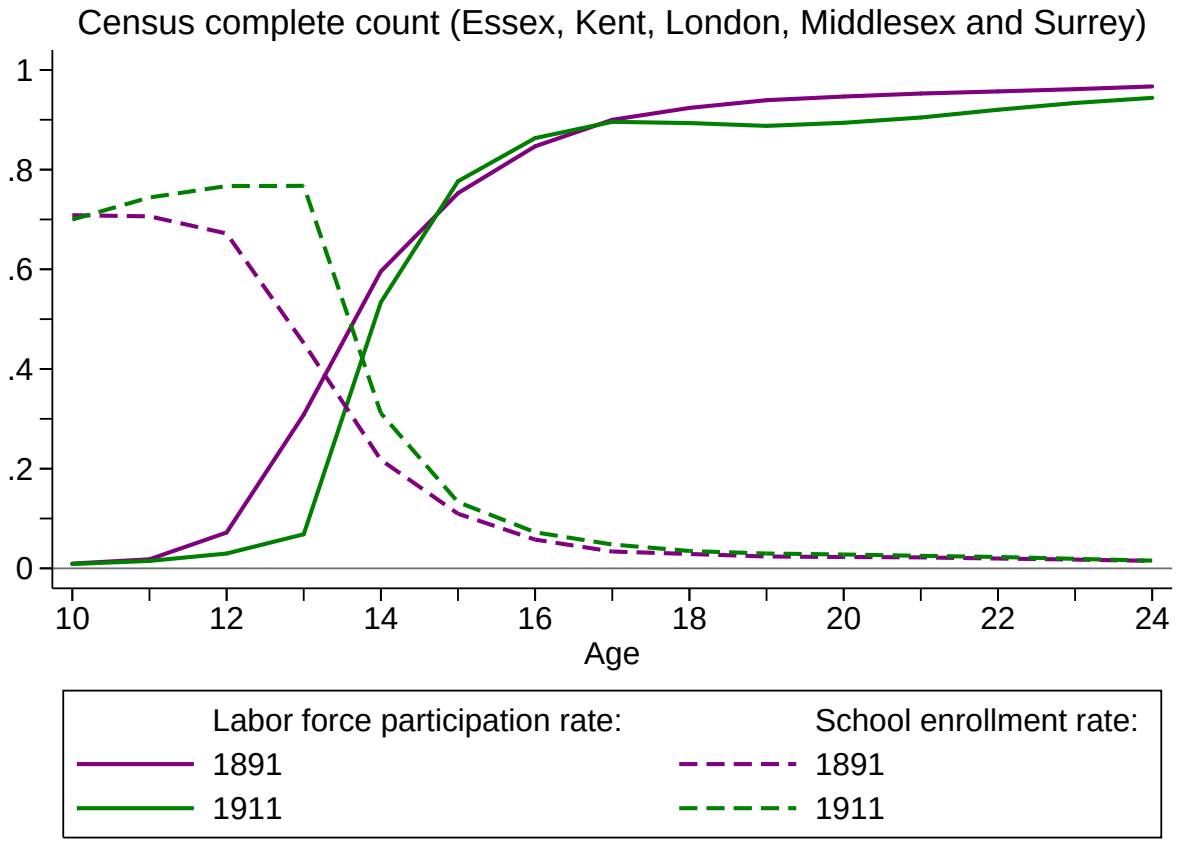
Figures

Figure 1: Distribution of patients' residences in final sample



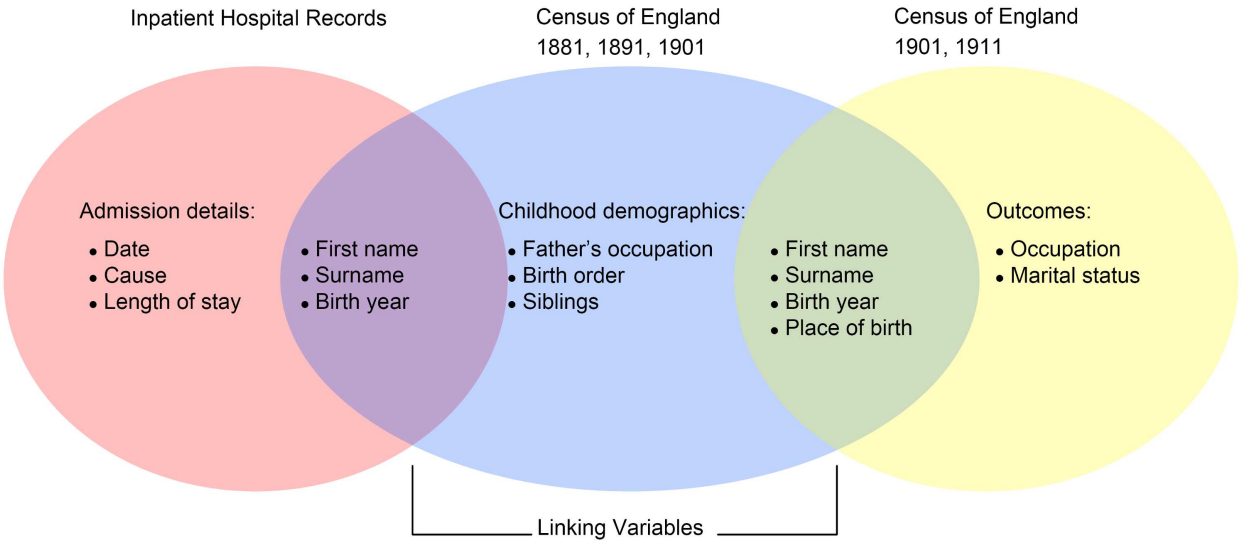
Notes: Figure 1 displays the share of patients residing in each registration district of London among all patients admitted to Barts and GOSH from the 1874 to 1890 birth cohorts between the ages of 0 and 11 in our final matched sample. Darker shades of blue denote a larger share of total admissions.

Figure 2: Labor force participation and school enrollment by age



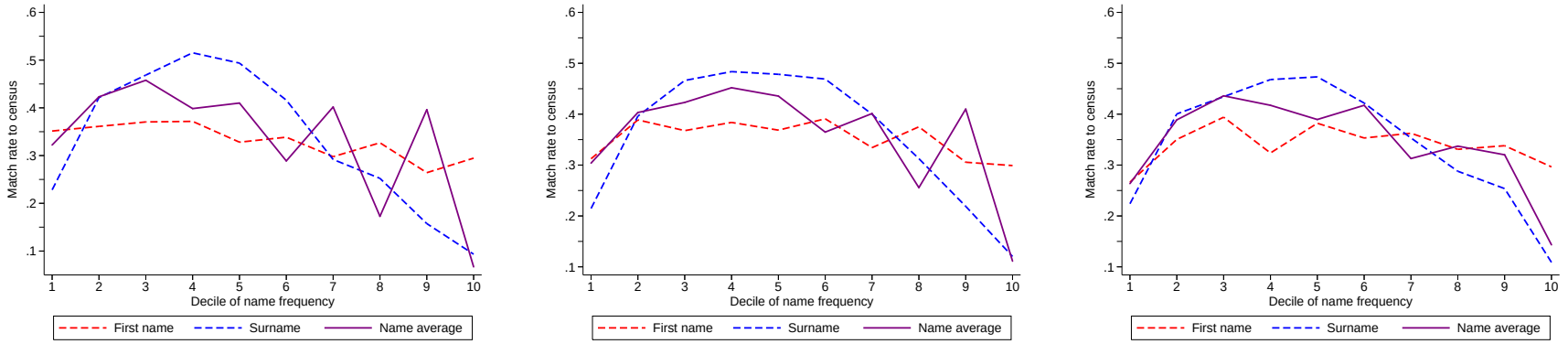
Notes: Figure 2 plots the labor force participation rate (solid lines) and school enrollment age (dashed line) by age (10 to 24) in the 1891 (purple lines) and 1911 (green lines) Population Censuses of England, for individuals residing in the counties of Essex, Kent, London, Middlesex, or Surrey. An individual is considered in school if the census records her occupation as "scholar," or in the labor force if any other occupation is recorded in the census.

Figure 3: Data sources and linking method for male sample



Notes: Figure 3 illustrates the data sources and variables used to link the hospital registers to census records for the sample of male patients. Each circle represents one stage in the linking procedure. The left-hand-side circle represents the hospital admission registers, the middle circle represents the Census of England in which the patient is located during childhood (1881, 1891, or 1901), and the circle on the right represents the census during adulthood (1901 or 1911). The variables listed in the overlapping portions of the Venn diagram are used to link the sources. Other variables of interest are listed in the non-overlapping portions of the circles.

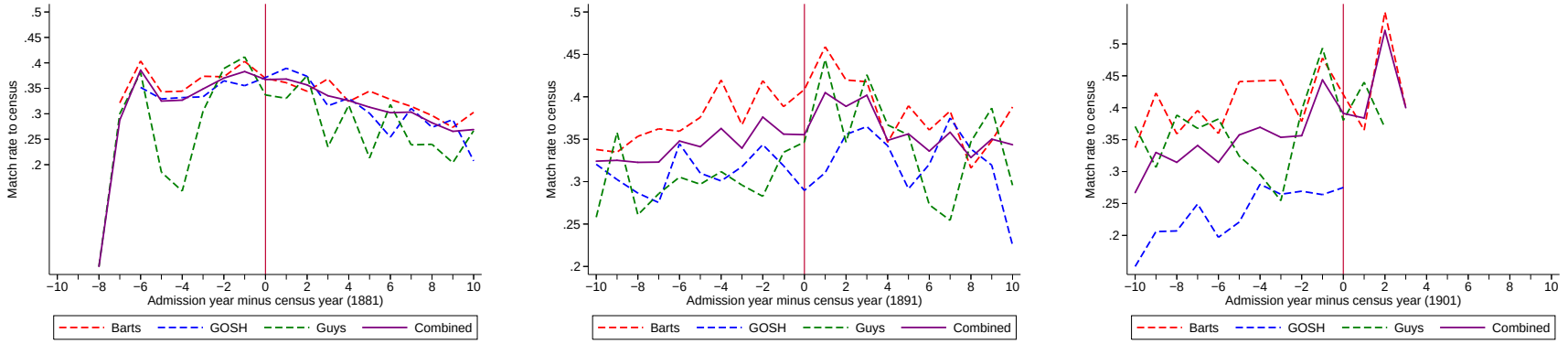
Figure 4: Match rates to childhood census by commonness of names



(a) 1881 (b) 1891 (c) 1901

Notes: Figure 4 plots the rate of unique matches from the hospital records to the census during childhood by decile of name frequency, separately for observations matched to the 1881, 1891 or 1901 censuses. Each figure contains separate plots for match rates of first names, surnames, and the average of the two names. The frequency of a name is based on the complete count of the 1881 Census of England.

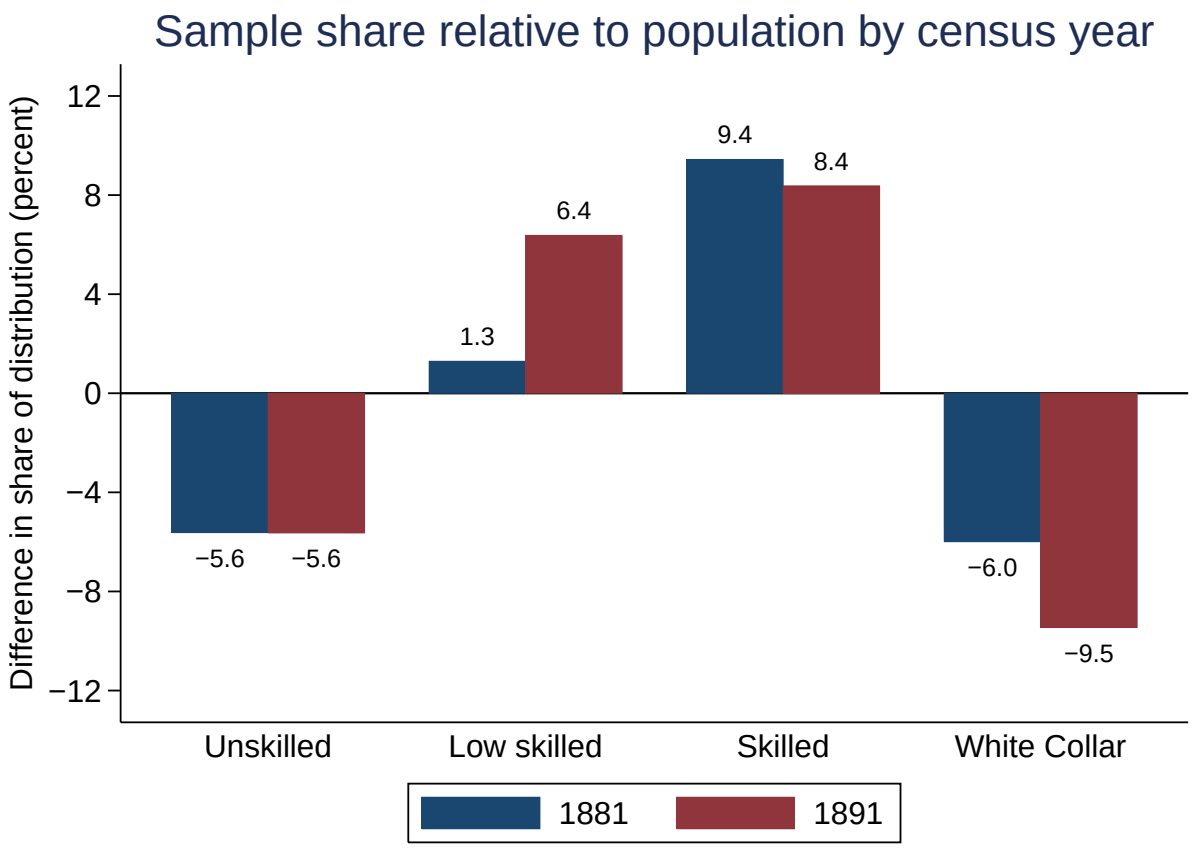
Figure 5: Match rates to childhood census by hospital admission year



(a) 1881 (b) 1891 (c) 1901

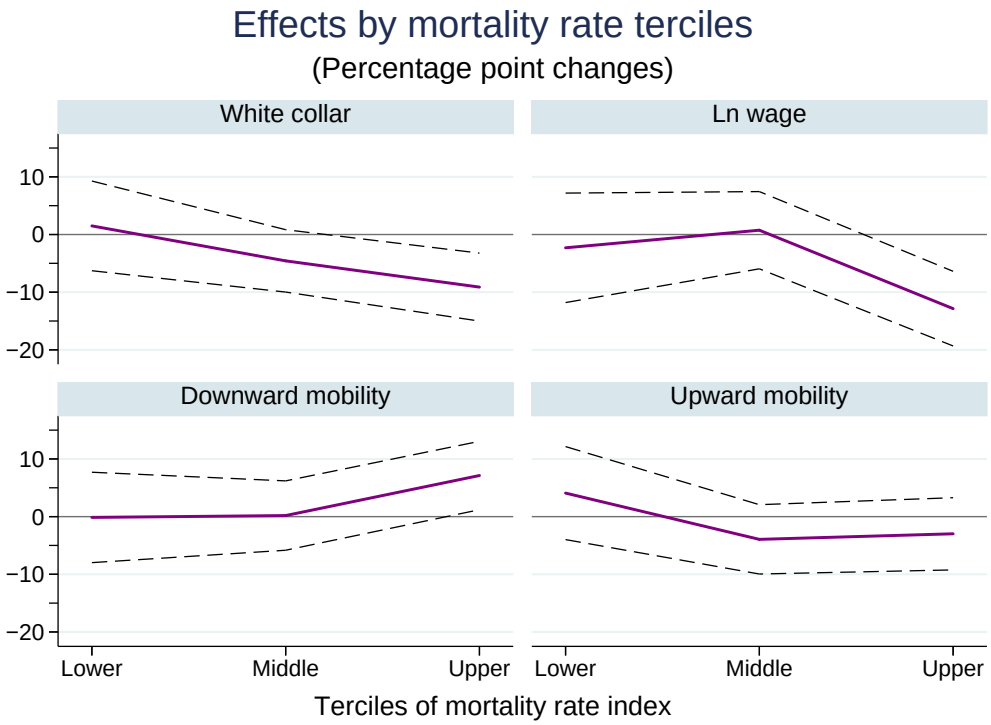
Notes: Figure 5 plots unique match rates from the hospital records to the census during childhood by year of admission to the hospital, separately for observations matched to the 1881, 1891 or 1901 censuses. Each figure contains separate plots of match rates for Barts and GOSH hospitals.

Figure 6: Occupational status of patients' fathers



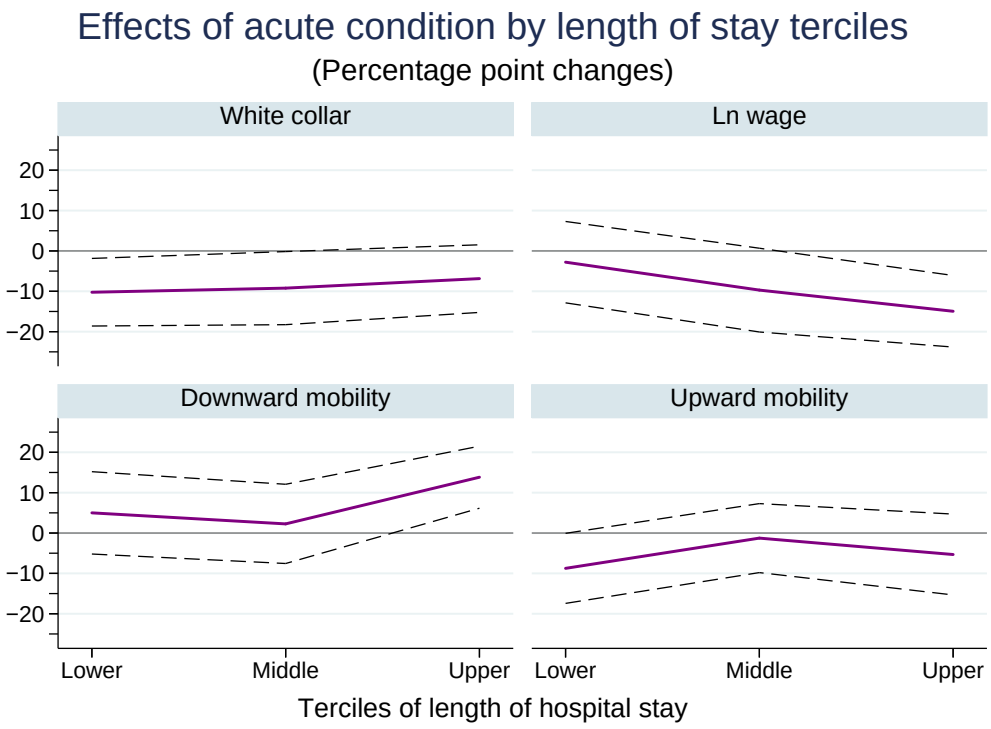
Notes: Figure 6 displays the distribution of occupations for fathers of hospitalized patients in our final regression sample relative to a comparable group in the population in the 1881 (blue) and 1891 (red) population censuses of England. The comparison group contains all fathers with at least one child between the ages of 0 and 12 present in the same household, and residing in the counties of London, Essex, Kent, Middlesex or Surrey.

Figure 7: Effects of admission conditions by terciles of in-hospital mortality rate



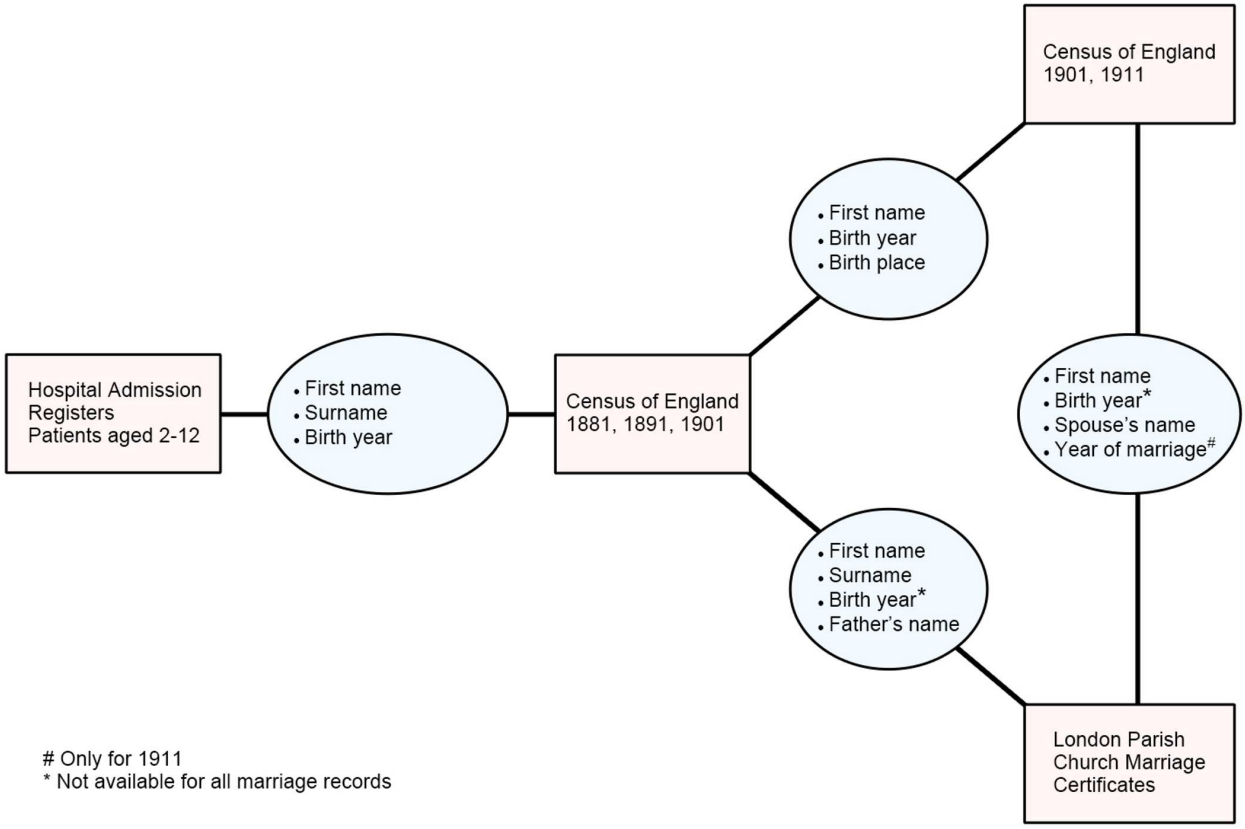
Notes: Figure 7 contains four plots of coefficients on the mortality rate index interacted with indicator variables for patients admitted with conditions in each tercile of the mortality index (lower, middle, or upper). Each plot shows coefficients from a regression with a different dependent variable: an indicator for entering a white collar occupation, the occupational log wage of an individual, an indicator for entering an occupation in a lower class than one's father, and an indicator for entering an occupation in a higher class than one's father (clockwise from top-left).

Figure 8: Effects of admission conditions by terciles of length of hospital stay



Notes: Figure 8 contains four plots of coefficients on an indicator for acute conditions interacted with indicator variables for patients admitted with conditions in each tercile of the length of stay in the hospital (lower, middle, or upper). Each plot shows coefficients from a regression with a different dependent variable: an indicator for entering a white collar occupation, the occupational log wage of an individual, an indicator for entering an occupation in a lower class than one's father, and an indicator for entering an occupation in a higher class than one's father (clockwise from top-left).

Figure 9: Data sources and linking method for female sample



Notes: Figure 9 illustrates the data sources and variables used to link the hospital registers to marriage and census records for the sample of female patients. The boxes represent the data sources at each stage of the linking procedure and the circles contain the variables used for each linkage. First, the hospital records are linked to a census during childhood (1881, 1891, or 1901) using full name and birth year. Second, the woman's maiden name, birth year and father's name are used to link the census record to a marriage certificate. Third, the woman's given name and birth year as well as her spouse's surname and year of marriage are used to link to a census during adulthood (1901 or 1911). Finally, the two census records are linked by verifying that the birth place matches across the sources.

Tables

Table 1: Number of beds and total admissions to voluntary hospitals in London.

Hospital (1894)	# of Beds	Inpatients	Outpatients	Inpatient %
Panel A: General hospitals				
Barts	675	6,474	159,802	4.05
Top-12 General	4,937	52,231	688,187	7.59
Barts share (%)	19.7	12.4	23.2	
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: Table 1 displays the total number of beds, inpatients, and outpatients admitted at voluntary hospital in London during 1894, as the well the share of inpatients among outpatients for different groups of hospitals. Panel A presents totals for general hospitals, including Barts hospital (one of two hospitals in our sample) and the largest twelve general hospitals, as well as Barts' share of each statistics among the top-12 hospitals. Panel B shows the same totals for children's hospitals, including GOSH (the second hospital in our sample) and the top-6 children's hospitals in London.

Table 2: Linkage rates between hospital registers and census records

	Hospital records to childhood census			Patients and siblings Childhood to adulthood census		
	(1) Total Admissions	(2) Match to Census	(3) Match with Siblings	(4) Match to Census	(5) Final Sample	(6) Total Admissions
Panel A: Fraction of observations remaining after each matching stage						
Barts	10,874	0.480	0.344	0.074	0.066	719
GOSH	7,924	0.462	0.345	0.068	0.065	519
	Match Failure	Multiple Matches	Unique Matches	Match Failure	Multiple Matches	Unique Matches
Panel B: Sources of sample attrition (fraction of observations)						
Barts	0.202	0.319	0.480	0.357	0.128	0.517
GOSH	0.300	0.238	0.462	0.310	0.173	0.537

Notes: Table 2 displays summary measures of the match rates between hospital records and censuses. Column (1) of Panel A shows the number of patients admitted to Barts and GOSH during our sample period from 1874 to 1901. Columns (2) to (5) show the percentage of admission that, respectively, are matched to the census during childhood, matched to a census during childhood along with a sibling, matched to a census during adulthood along with a sibling, and appear in the final sample. Column (6) shows the total admissions in the final sample. In Panel B, columns (1) to (3) show the distribution of observations that result in match failure, multiple matches and unique matches for each hospital, while columns (4) to (6) show the same for linking the censuses during childhood and adulthood.

Table 3: Hospital admission characteristics.

	(1) All Admissions	(2) Matched to Census	(3) Final Sample
Age on admission (0-11 years)	5.42	**5.77**	**5.75**
Infant admission (age 0-1)	0.13	**0.09**	0.09
Length of admission (days)	30.11	30.22	30.23
Multiple admissions	0.17	0.17	0.16
Winter admissions	0.48	0.48	0.47
Resident outside Greater London	0.06	**0.00**	**0.00**
Death in hospital	0.10	**0.00**	**0.00**
High SES father		0.43	0.46
Sibship size		3.77	4.86
Observations	18,798	8,876	1,238

Notes: Table 3 displays summary statistics for all hospital admissions (column (1)), hospital patients matched to a census during childhood (column (2)), and observations in the final estimation sample (column (3)).

Table 4: Effects of hospital admission and in-hospital mortality rate.

	Individual occupational success				Mobility	
	(1) White collar	(2) Skilled +	(3) Semi-skilled +	(4) Ln wage	(5) Class ↘	(6) Class ↗
Panel A: Effects of hospital admission						
Patient	−0.039**	−0.043**	−0.021	−0.042*	0.022	−0.010
	(0.018)	(0.021)	(0.016)	(0.022)	(0.019)	(0.019)
Panel B: Effects of mortality rate for admitted condition						
Mortality rate	−0.316***	−0.264**	−0.167*	−0.361***	0.246**	−0.189*
	(0.102)	(0.129)	(0.101)	(0.117)	(0.119)	(0.109)
Mean of Y	0.269	0.528	0.843	4.632	0.301	0.351
Observations	2184	2184	2184	1624	1802	1802

Notes: Each column and panel displays coefficients from separate OLS regressions. In columns (1) to (3) the dependent variables are indicators for individual occupational success which take the value of one if an individual enters a white collar occupation; a white collar or skilled occupation; and a white collar, skilled, or semi-skilled occupation, respectively. In column (4), the dependent variable is the occupational log wage assigned to the individual's occupational title as an adult. Columns (5) and (6) display coefficients from regressions with measures of intergenerational mobility as outcomes. The dependent variables equal one if an individual enters an occupation lower in class than his father, or higher in the class than his father, respectively. Columns (4) to (6) have fewer observations due to missing values of the occupational wage and the father's occupation. Panel A presents coefficients on an indicator for whether an individual appeared in the hospital admission records in our sample, while Panel B display coefficients on the in-hospital mortality rate of the admitted condition. See Appendix 9 for details on how this variable is constructed. All regressions include household, birth year, and birth order fixed effects, as well as indicators for the younger sibling in a sibling pair and the census year in which the outcome variable is observed. Standard errors are clustered at the household level.

Table 5: Effects of admission severity and treatment by doctor.

	Individual occupational success				Mobility	
	(1) White collar	(2) Skilled +	(3) Semi-skilled +	(4) Ln wage	(5) Class ↘	(6) Class ↗
Panel A: Effects of admission for acute condition						
Acute	−0.090***	−0.094***	−0.023	−0.092***	0.069**	−0.046*
	(0.026)	(0.029)	(0.024)	(0.029)	(0.028)	(0.027)
Not acute	0.009	0.006	−0.017	0.005	−0.009	0.023
	(0.022)	(0.026)	(0.019)	(0.028)	(0.024)	(0.024)
P-value	0.004	0.010	0.841	0.016	0.034	0.058
Panel B: Effects of treatment by doctor						
Doctor	−0.073***	−0.093***	−0.057**	−0.107***	0.066**	−0.052
	(0.028)	(0.034)	(0.028)	(0.034)	(0.032)	(0.032)
No doctor	−0.013	−0.008	−0.005	−0.019	−0.003	0.020
	(0.021)	(0.024)	(0.017)	(0.026)	(0.022)	(0.022)
P-value	0.086	0.037	0.113	0.040	0.081	0.057
Mean of Y	0.269	0.528	0.843	4.632	0.301	0.351
Observations	2184	2184	2184	1624	1802	1802

Notes: Each column and panel displays coefficients from separate OLS regressions. See Table 4 for a description of the dependent variables in each column. Panel A displays coefficients on two indicators that take the value of one if an individual was admitted to a hospital in our sample with an acute or non-acute conditions, respectively. Panel B presents coefficients on two indicator variables that equal one if an individual was treated by a doctor, on one hand, or treated by a surgeon or medical assistant, on the other. All regressions include household, birth year, and birth order fixed effects, as well as indicators for the younger sibling in a sibling pair and the census year in which the outcome variable is observed. Standard errors are clustered at the household level.

Table 6: Ordered probit: marginal effects of severe conditions.

	(1) White collar	(2) Skilled	(3) Semi-skilled	(4) Unskilled
Panel A: Effects of in-hospital mortality rate				
Mortality rate	-0.124^{**}	-0.049	0.044^{***}	0.130
	(0.053)	(0.034)	(0.009)	(0.080)
Panel B: Effects of admission with acute condition				
Acute	-0.039^{**}	-0.010^{*}	0.018^{**}	0.032^{**}
	(0.017)	(0.005)	(0.007)	(0.015)
Non-acute	-0.014	-0.004	0.007	0.016
	(0.016)	(0.004)	(0.007)	0.016)
Panel C: Effects of treatment by doctor				
Doctor	-0.058^{***}	-0.016^{**}	0.025^{***}	0.049^{***}
	(0.018)	(0.007)	(0.007)	(0.017)
No doctor	-0.005	-0.002	0.002	0.004
	(0.015)	(0.004)	(0.007)	0.012

Notes: Table 6 presents coefficients from ordered probit regressions. The dependent variables are indicator variables for individual occupational success and correspond to the dependent variables in columns (1) to (4) of Table 4. Each Panel displays the marginal effects of different measures of severity: the in-hospital mortality rate of the admitted condition (Panel A), acute conditions (Panel B), and treatment by a doctor (Panel C). All regressions include birth year, and birth order fixed effects, as well as indicators for the younger sibling in a sibling pair and the census year in which the outcome variable is observed.

Table 7: Benchmark for effects on occupational status.

	(1) White collar	(2) Skilled+	(3) Ln wage
Panel A: Intergenerational occupational elasticities			
Father's status (γ)	0.228***	0.161***	0.283***
	(0.036)	(0.027)	(0.050)
Panel B: Scaled effects			
Patient	0.170	0.264	0.149
Mortality rate	0.137	0.163	0.124
Acute condition	0.393	0.583	0.327
Treated by doctor	0.319	0.580	0.378

Notes: Panel A presents estimates of intergenerational occupational elasticities. In column (1) the dependent variable (independent regressor of interest) is an indicator for whether the son (father) worked in a white collar occupation, in column (2) it is an indicator for whether the son (father) worked in a white collar occupation, and in column (3) it is the occupational log wage of the son (father). In Panel B, columns (1) and (2) present the coefficients from Tables 4 and 5 scaled by the estimates in Panel A. In column (3), the coefficients are scaled by a one-standard-deviation change in the father's occupational log wage.

Table 8: Benchmark for effects on intergenerational mobility.

	(1) Linked sample	(2)	(3) England, 1881-1901†	(4)
	Class ↘	Class ↗	Class ↘	Class ↗
Panel A: Estimates of mobility rates				
Sample mean	0.301	0.351	0.215	0.268
Panel B: Effects of poor health scaled by sample mobility rates				
Mortality rate	0.082	0.054	0.115	0.071
Acute condition	0.231	0.131	0.323	0.171
Treated by doctor	0.219	0.149	0.306	0.194

Notes: Panel A presents means of the dependent variable in the estimation sample (columns (1) and (2)) and a representative sample for England linking the 1881 and 1901 censuses (columns (3) and (4)). The dependent variables are indicators for an individual who enters a lower occupational class than his father (columns (1) and (3)) or a higher occupational class (columns (2) and (4)). Panel B presents coefficients from Tables 4 and 5 scaled by the sample means in Panel A. † Source: Long (2013).

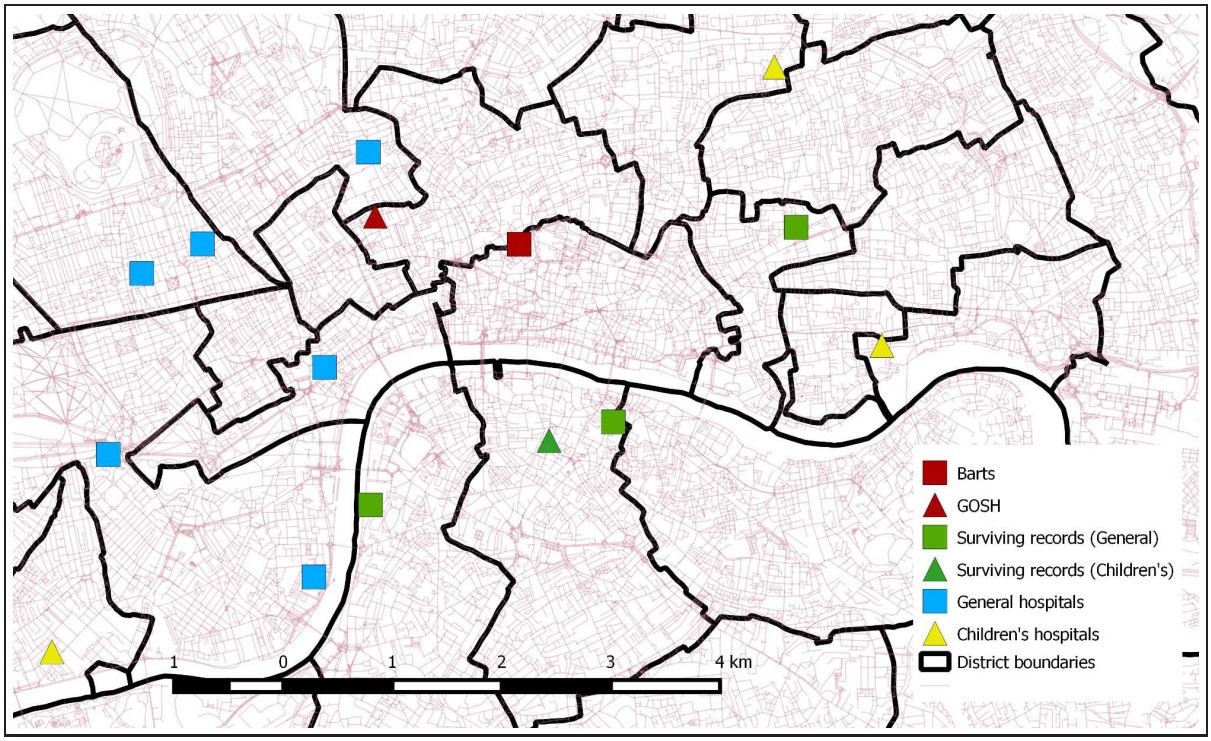
Table 9: Main results: Marriage

	(1) Males	(2) Females
Panel A: Effect of hospital admission		
Patient	−0.009	−0.086***
	(0.018)	(0.029)
Panel B: Effect of admission for acute condition		
Acute condition	−0.031	−0.104**
	(0.025)	(0.043)
Non-acute condition	0.005	−0.059*
	(0.022)	(0.033)
Mean of Y	0.516	0.337
# FEs	1,080	369
Observations	2,435	806

Notes: Table 9 displays coefficients from regressions in which the dependent variable is an indicator that takes the value of one if an individual is married at the time of census enumeration during adulthood. Columns (1) and (2) display results from the samples of male and female patients, respectively. Panel A displays coefficients on the indicator for hospitalization, while Panel B shows the coefficients on acute and non-acute conditions. All regressions include household, birth year, and birth order fixed effects, as well as indicators for the younger sibling in a sibling pair and the census year in which the outcome variable is observed. Standard errors are clustered at the household level.

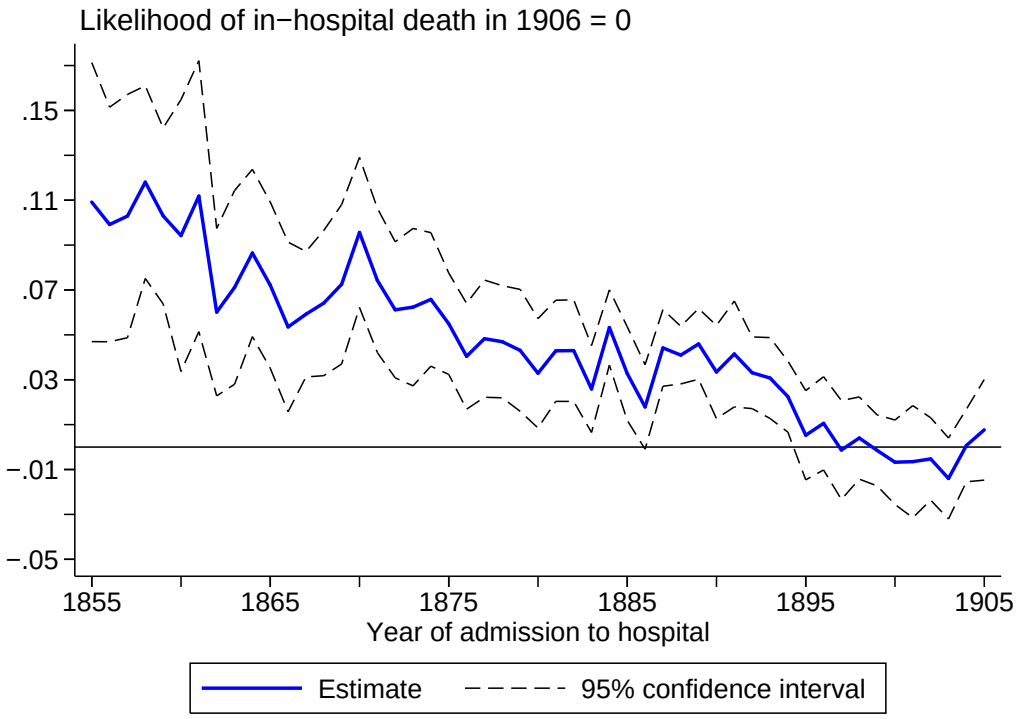
Appendix Figures

Figure A1: General and children's hospital in nineteenth century London



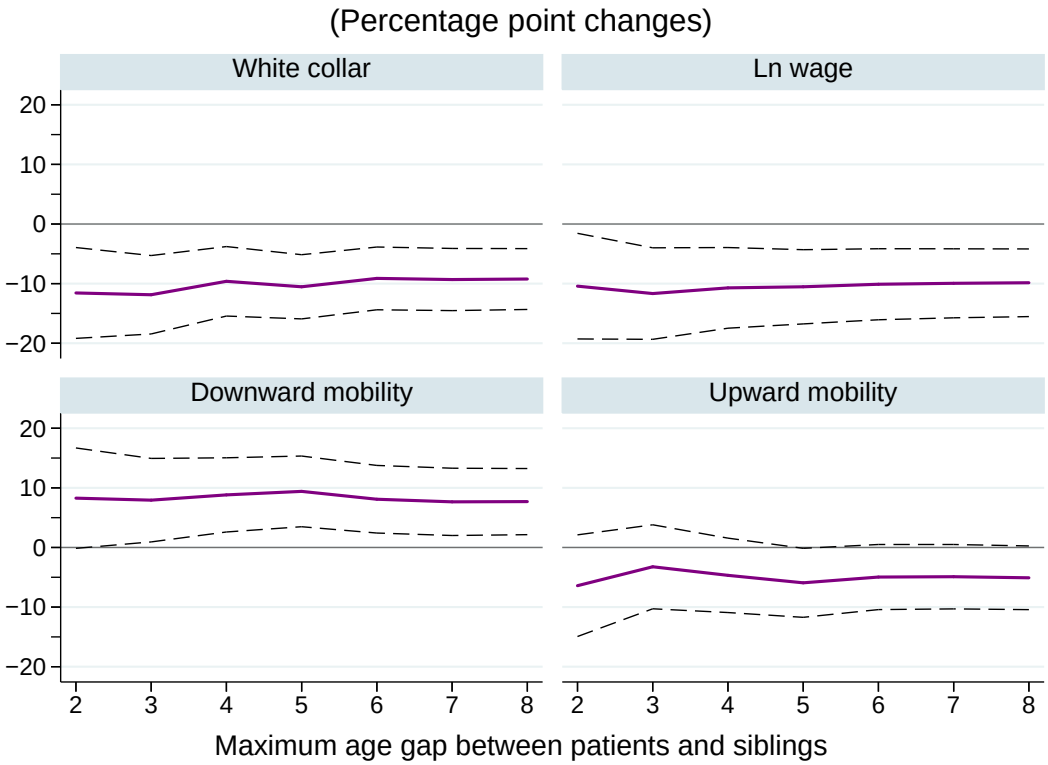
Notes: Figure A1 displays the locations of prominent general and children's hospitals in London at the end of the nineteenth century (blue and yellow, respectively), the locations of hospitals with surviving inpatient records (green), and the hospitals in our sample (red).

Figure A2: Probability of death in hospital, 1855-1905



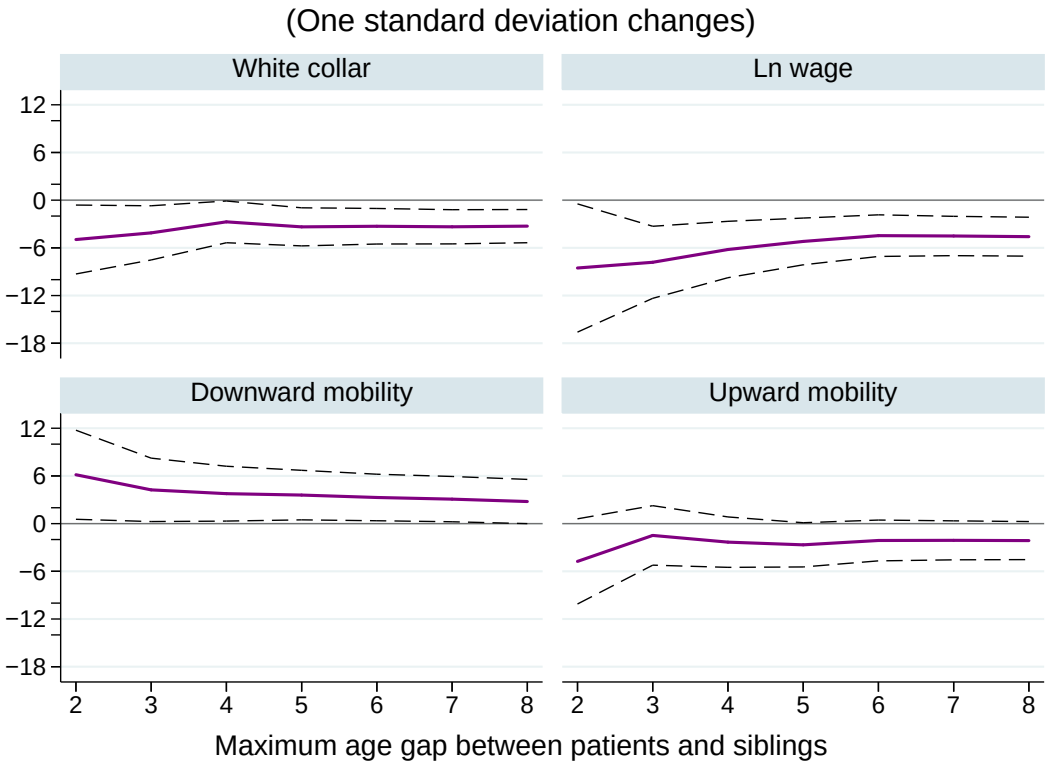
Notes: Figure A2 plots coefficients on year-of-admission fixed effects and 95% confidence intervals from a district-of-residence-FE model in which the dependent variable is an indicator for admissions leading to death in the hospital. The independent variables are indicators for patients residing outside London, repeated admissions, acute conditions, and injuries as the cause of admission, as well as the length of stay in the hospital (in days) and a linear time trend. The regression also controls for age-at-admission FE, month-of-admission FE, hospital FE, and disease-category FE. The 95% confidence intervals are computed using heteroskedasticity robust standard errors clustered at the registration-district level.

Figure A3: Effects of acute conditions by maximum age gap between patients and siblings



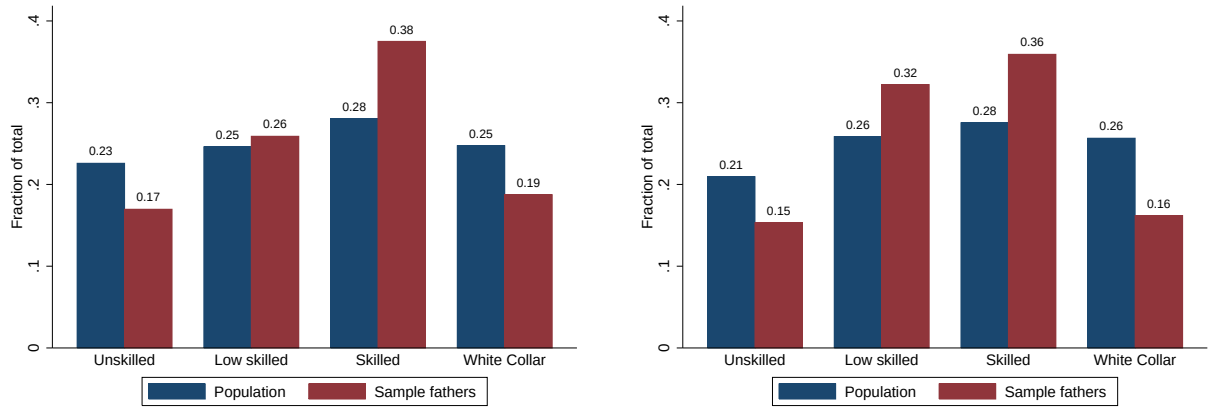
Notes: Figure A3 contains four plots of coefficients from regressions with different dependent variables: an indicator for entering a white collar occupation, the occupational log wage of an individual, an indicator for entering an occupation in a lower class than one's father, and an indicator for entering an occupation in a higher class than one's father (clockwise from top-left). Each plot shows coefficients on acute conditions from separate regressions in which the maximum age gap between patients and siblings varies from 2 to 8 years.

Figure A4: Effects of mortality index by maximum age gap between patients and siblings



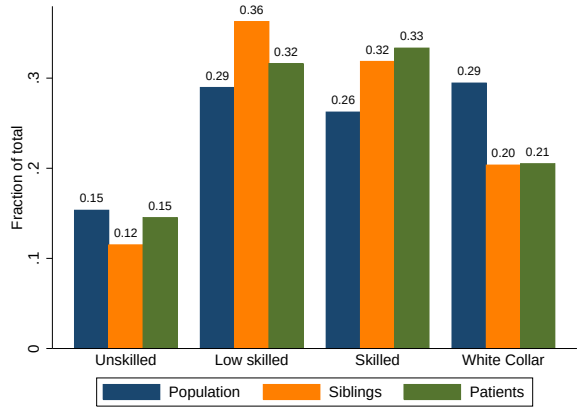
Notes: Figure A4 contains four plots of coefficients from regressions with different dependent variables: an indicator for entering a white collar occupation, the occupational log wage of an individual, an indicator for entering an occupation in a lower class than one's father, and an indicator for entering an occupation in a higher class than one's father (clockwise from top-left). Each plot shows coefficients on the mortality rate index from separate regressions in which the maximum age gap between patients and siblings varies from 2 to 8 years.

Figure A5: Occupational distribution of fathers and sons in sample and population

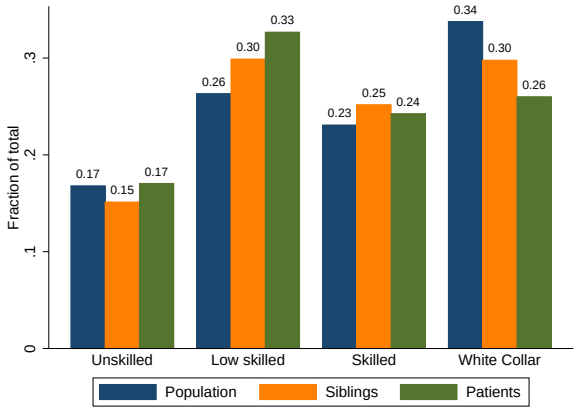


(a) 1881 Census of England

(b) 1891 Census of England



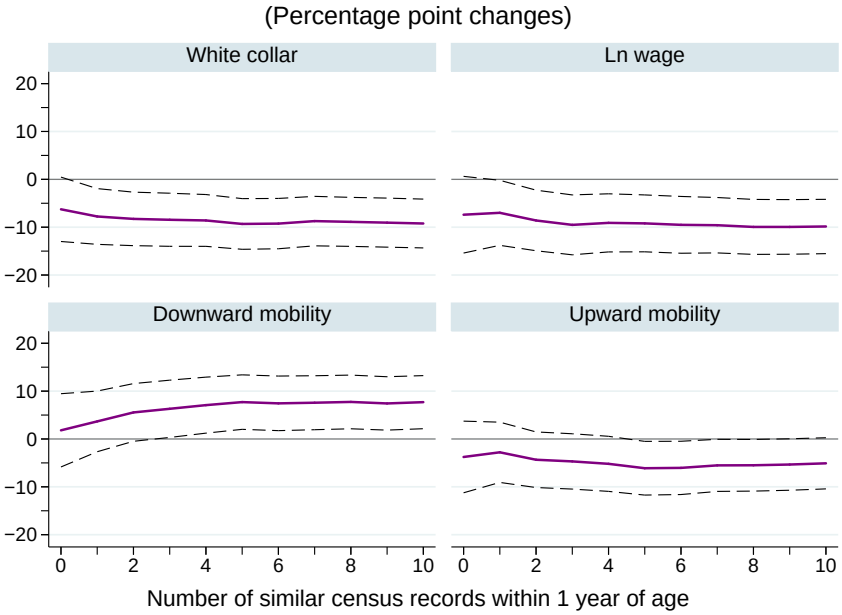
(c) 1901 Census of England



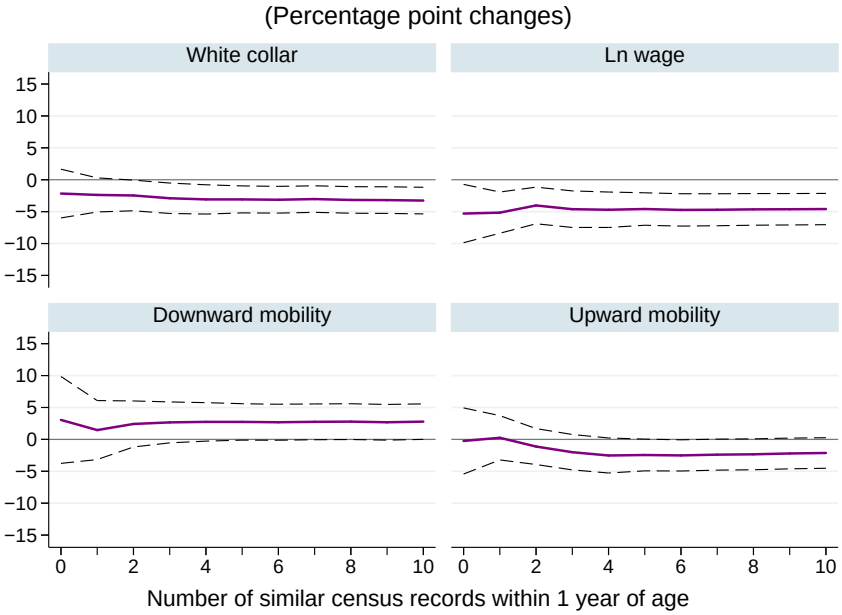
(d) 1911 Census of England

Notes: Figures A5a and A5b plot the distribution of occupational classes (unskilled, semi-skilled, skilled or white collar) for the population of household heads in London and surrounding counties, and for fathers of patients in our sample, in the 1881 and 1891 censuses, respectively. Figures A5c and A5d plot the distribution of occupational classes for the population of males aged 18 to 37, in addition to patients and siblings in our sample, in the 1901 and 1911 censuses, respectively.

Figure A6: Effects by number of similar census records within 1 year of age



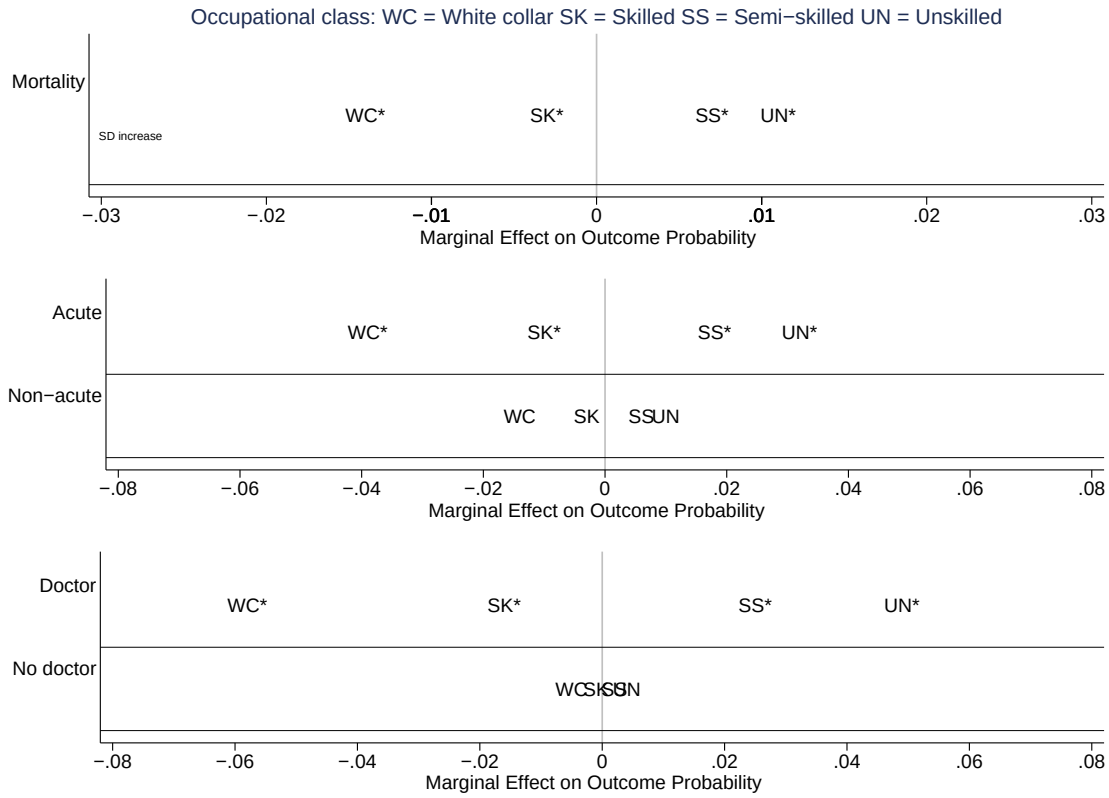
(a) Effects of acute conditions on outcomes



(b) Effects of mortality index on outcomes

Notes: Figures A6a and A6b each contain four plots of coefficients on acute conditions and the mortality index, respectively, from regressions with different dependent variables: an indicator for entering a white collar occupation, the occupational log wage of an individual, an indicator for entering an occupation in a lower class than one's father, and an indicator for entering an occupation in a higher class than one's father (clockwise from top-left). Each plot shows coefficients on acute conditions (Figure A6a) or the mortality index (Figure A6b)from separate regressions in which the number of similar records within one year of age of the matched record in the census varies between 0 and 10.

Figure A7: Magnitudes of marginal effects in ordered probit model



Notes: Figure A7 shows the magnitudes of the marginal effects of the mortality index, acute and non-acute conditions, and treatment by a doctor or surgeon from ordered probit models.

Appendix Tables

Table A1: Common occupation titles by occupational class.

White collar	Skilled	Low-skilled	Unskilled
Clerk	Carpenter	Carman	General laborer
Police constable	Bricklayer	House painter	Laborer
Bank clerk	Butcher	Postman	Farm laborer
Railway clerk	Compositor	Barman	General domestic
Shop assistant	Plumber	Porter	Warehouseman

Notes: Table A1 lists the five most common occupations in each of four occupational classes for the final sample of patients and siblings used in the main regressions. Column (1) combines professional, managerial and clerical occupations (Classes 1 and 2 in in the seven category HISCLASS scheme) into a white collar class, Column (2) subsumes farmers into skilled workers (HISCLASS 3 and 4), Column (3) displays low-skilled workers (HISCLASS 5), and Column (4) combines unskilled workers as well as low and unskilled farm workers (HISCLASS 6 and 7)

Table A2: Common conditions by acute status and type of medical provider.

	(1) All Admissions		(2) Regression Sample	
	Disease	(%)	Disease	(%)
	Panel A: Acute conditions			
1	Abscess	11.05	Abscess	14.81
2	Diphtheria & sequela	9.51	Pneumonia	10.49
3	Bronchitis	9.12	Bronchitis	7.20
4	Pneumonia	7.56	Diphtheria & sequela	7.00
5	Tuberculosis (Phthisis)	5.79	Tuberculosis (Phthisis)	2.88
	Panel B: Non-acute conditions			
1	Fractured bone	6.51	Fractured bone	6.78
2	Disease hip	4.12	Disease hip	5.85
3	Chorea	3.59	Chorea	4.52
4	Phimosis	2.83	Phimosis	3.99
5	Disease knee	2.66	Disease knee	3.46
	Panel C: Treatment by doctors			
1	Chorea	6.73	Pneumonia	8.15
2	Pneumonia	5.19	Chorea	6.00
3	Bronchitis	3.69	Diphtheria & sequela	5.28
4	Morbus cordis	2.75	Bronchitis	4.80
5	Empyema	2.58	Debility	2.16
	Panel D: Treatment by surgeons and assistants			
1	Disease knee	5.08	Phimosis	7.51
2	Disease hip	4.93	Disease hip	6.97
3	Phimosis	4.70	Disease knee	5.63
4	Talipes	4.16	Abscess	4.02
5	Concussion	3.46	Fractured femur	3.75

Notes: Table A2 displays the most common conditions among all admissions in the hospital records data base (column (1)) and among admissions in the final estimation sample (column (2)). The five most frequently observed causes of admission are presented for acute conditions (Panel A), non-acute conditions (Panel B), cases treated by a doctor (Panel C), and cases treated by surgeons or medical assistants (Panel D).

Table A3: Robustness to using Armstrong classification instead of HISCO

	Individual occupational success				Mobility	
	(1) White collar	(2) Skilled +	(3) Semi-skilled +	(4) Ln wage	(5) Class ↘	(6) Class ↗
Panel A: Effects of hospital admission						
Patient	−0.043**	−0.057**	−0.051**	−0.042*	0.022	−0.010
	(0.017)	(0.024)	(0.020)	(0.022)	(0.019)	(0.019)
Panel B: Effects of mortality rate for admitted condition						
Mortality rate	−0.354***	−0.395***	−0.255*	−0.361***	0.246**	−0.189*
	(0.101)	(0.148)	(0.135)	(0.117)	(0.119)	(0.109)
Mean of Y	0.189	0.635	0.802	4.632	0.301	0.351
Observations	1624	1624	1624	1624	1802	1802

Notes: Each column and panel displays coefficients from separate OLS regressions. Table A3 follows the format and specifications presented in Table 4, but uses the Armstrong (1972) classification scheme to assign occupational titles to occupational classes, instead of the HISCO scheme (Leeuwen and Maas, 2011) used in the main specifications.

Table A4: Robustness to using Armstrong classification instead of HISCO

	Individual occupational success				Mobility	
	(1) White collar	(2) Skilled +	(3) Semi-skilled +	(4) Ln wage	(5) Class ↘	(6) Class ↗
Panel A: Effects of admission for acute condition						
Acute	−0.080***	−0.098***	−0.043	−0.092***	0.069**	−0.046*
	(0.024)	(0.034)	(0.030)	(0.029)	(0.028)	(0.027)
Not acute	−0.009	−0.017	−0.046*	0.005	−0.009	0.023
	(0.021)	(0.029)	(0.025)	(0.028)	(0.024)	(0.024)
P-value	0.029	0.073	0.941	0.016	0.034	0.058
Panel B: Effects of treatment by doctor						
Doctor	−0.068**	−0.151***	−0.086***	−0.107***	0.066**	−0.052
	(0.028)	(0.039)	(0.033)	(0.034)	(0.032)	(0.032)
No doctor	−0.033	−0.004	−0.026	−0.019	−0.003	0.020
	(0.020)	(0.027)	(0.023)	(0.026)	(0.022)	(0.022)
P-value	0.291	0.002	0.134	0.040	0.081	0.057
Mean of Y	0.189	0.635	0.802	4.632	0.301	0.351
Observations	1624	1624	1624	1624	1802	1802

Notes: Each column and panel displays coefficients from separate OLS regressions. Table A4 follows the format and specifications presented in Table 5, but uses the Armstrong (1972) classification scheme to assign occupational titles to occupational classes, instead of the HISCO scheme (Leeuwen and Maas, 2011) used in the main specifications.

Table A5: Effects of hospital admission and in-hospital mortality rate. OLS.

	Individual occupational success				Mobility	
	(1) White collar	(2) Skilled +	(3) Semi-skilled +	(4) Ln wage	(5) Class ↘	(6) Class ↗
Panel A: Effects of hospital admission						
Patient	−0.034*	−0.037*	−0.018	−0.033	0.022	−0.004
	(0.018)	(0.020)	(0.015)	(0.022)	(0.019)	(0.019)
Panel B: Effects of mortality rate for admitted condition						
Mortality rate	−0.207**	−0.142	−0.079	−0.286**	0.110	−0.090
	(0.092)	(0.107)	(0.087)	(0.114)	(0.106)	(0.105)
Mean of Y	0.269	0.528	0.843	4.632	0.301	0.351
Observations	2184	2184	2184	1624	1802	1802

Notes: Each column and panel displays coefficients from separate OLS regressions. Table A5 follows the format presented in Table 4, but does not include household fixed effects in the specifications.

Table A6: Effects of admission severity and treatment by doctor. OLS.

	Individual occupational success				Mobility	
	(1) White collar	(2) Skilled +	(3) Semi-skilled +	(4) Ln wage	(5) Class ↘	(6) Class ↗
Panel A: Effects of admission for acute condition						
Acute	−0.052**	−0.051**	−0.019	−0.048*	0.028	0.008
	(0.023)	(0.026)	(0.020)	(0.028)	(0.026)	(0.026)
Not acute	−0.013	−0.015	−0.012	−0.019	0.015	−0.004
	(0.021)	(0.023)	(0.017)	(0.026)	(0.022)	(0.023)
P-value	0.156	0.250	0.779	0.414	0.696	0.719
Panel B: Effects of treatment by doctor						
Doctor	−0.060**	−0.072**	−0.048**	−0.062**	0.071**	−0.041
	(0.024)	(0.029)	(0.023)	(0.031)	(0.029)	(0.029)
No doctor	−0.008	−0.008	0.001	−0.026	−0.013	0.029
	(0.020)	(0.022)	(0.016)	(0.024)	(0.021)	(0.022)
P-value	0.060	0.051	0.056	0.320	0.014	0.038
Mean of Y	0.269	0.528	0.843	4.632	0.301	0.351
Observations	2184	2184	2184	1624	1802	1802

Notes: Each column and panel displays coefficients from separate OLS regressions. Table A6 follows the format and specifications presented in Table 5, but does not include household fixed effects in the specifications.

Table A7: Intergenerational mobility matrix in sample and population.

	Father's occupational class					Total	*N*
	Professional	Clerical	Skilled	Semi-skilled	Unskilled		
Panel A: Son's class in 1901/1911 (patients and siblings)							
Professional	21.6	17.9	12.7	14.0	8.0	13.6	245
Clerical	16.1	34.8	9.0	12.5	10.1	12.5	225
Skilled	17.9	17.0	36.7	21.3	22.6	26.9	484
Semi-skilled	34.6	20.5	28.5	36.9	35.9	32.3	582
Unskilled	9.9	9.8	13.1	15.4	23.4	14.8	266
N	162	112	694	586	248		
Panel B: Sons aged 30 to 39 in 1901 (Long, 2013)							
Professional	22.2	5.5	2.2	1.3	0.2	2.5	102
Clerical	27.8	31.0	11.5	6.3	7.7	12.4	505
Skilled	33.3	45.6	69.2	42.9	54.2	58.9	2,398
Semi-skilled	9.7	10.3	6.8	31.1	11.1	12.7	519
Unskilled	6.9	7.7	10.2	18.4	26.8	13.4	547
N	72	439	2,266	842	452		

Notes: Table A7 displays the transition matrix for intergenerational mobility between fathers and sons. The white collar occupational class in the main regression specifications has been separated into professional and clerical categories. Panel A shows the occupational transition probabilities for patients and siblings observed in the 1901 or 1911 census, in comparison to their fathers. Panel B presents occupational transition probabilities for a representative sample of sons aged 30 to 39 in 1901 across England from Long (2013).

Table A8: Method for computing in-hospital mortality rates.

Cause of admission: Abd. pain, ?Enteric fever

	(1) Sample	(2)	(3) Mortality rate	(4)
Component	Frequency	Sample	Period	Overall
Abdomen	162	0.148	0.127	0.142
Enteric fever	101	0.109	0.111	0.101
Pain	63	0.127	0.099	0.115

Notes: Table A8 presents an example of computing the in-hospital mortality rate for the condition "Abd. pain, ?Enteric fever." Column (1) gives the number of observations in the data that contain each component of the cause of admission string. Columns (2) to (4) present the mortality rates for each component calculated over the all observations from the 1874 to 1890 birth cohorts admitted between 1874 and 1901, all cohorts admitted between 1874 and 1891, and all patients admitted between 1870 and 1901, respectively.

Table A9: Sensitivity to selective mortality

	Individual occupational success				Mobility	
	(1) White collar	(2) Skilled +	(3) Semi-skilled +	(4) Ln wage	(5) Class ↘	(6) Class ↗
Panel A: Removing high-mortality conditions (Top 5%)						
Mortality rate	−0.384***	−0.410***	−0.179	−0.399***	0.287**	−0.176
	(0.133)	(0.153)	(0.113)	(0.145)	(0.130)	(0.149)
Mean of Y	0.268	0.528	0.844	4.636	0.300	0.351
Observations	2120	2120	2120	1578	1748	1748
Panel B: Removing contagious diseases						
Mortality rate	−0.472***	−0.391**	−0.213	−0.300*	0.307*	−0.192
	(0.146)	(0.174)	(0.142)	(0.177)	(0.173)	(0.165)
Mean of Y	0.271	0.530	0.842	4.630	0.302	0.347
Observations	1850	1850	1850	1376	1520	1520
Panel C: Removing patients admitted as infants (Age 0 to 1)						
Mortality rate	−0.326***	−0.290**	−0.191*	−0.413***	0.271**	−0.182
	(0.109)	(0.137)	(0.109)	(0.123)	(0.123)	(0.115)
Mean of Y	0.268	0.528	0.847	4.637	0.302	0.352
Observations	1962	1962	1962	1468	1618	1618
Panel D: Removing patients admitted multiple times						
Mortality rate	−0.390***	−0.368***	−0.240**	−0.351***	0.292**	−0.242*
	(0.117)	(0.137)	(0.115)	(0.130)	(0.132)	(0.124)
Mean of Y	0.269	0.526	0.842	4.629	0.298	0.352
Observations	1920	1920	1920	1416	1596	1596

Notes: Each column and panel displays coefficients from separate OLS regressions. See Table 4 for a description of the dependent variables in each column. Each Panel imposes restrictions that exclude certain observations from the main estimation sample. Panel A removes households with patients admitted for conditions with mortality rates in the top-5 percent, Panel B removes households with patients admitted for contagious diseases, Panel C removes households with patients admitted as infants (age 0 to 1), and Panel D removes households with patients admitted multiple times. All regressions include household, birth year, and birth order fixed effects, as well as indicators for the younger sibling in a sibling pair and the census year in which the outcome variable is observed. Standard errors are clustered at the household level.

Table A10: Sensitivity to data used to calculate in-hospital mortality rate

	Individual occupational success				Mobility	
	(1) White collar	(2) Skilled +	(3) Semi-skilled +	(4) Ln wage	(5) Class ↘	(6) Class ↗
Panel A: Mortality rate calculated using cohorts and admission years in sample						
Mortality rate	−0.316***	−0.264**	−0.167*	−0.361***	0.246**	−0.189*
	(0.102)	(0.129)	(0.101)	(0.117)	(0.119)	(0.109)
Panel B: Mortality rate calculated using sample-period admission data						
Mortality Rate	−0.302***	−0.264**	−0.183*	−0.323**	0.185	−0.156
	(0.104)	(0.128)	(0.098)	(0.127)	(0.120)	(0.112)
Panel C: Mortality rate calculated using all admission data						
Mortality Rate	−0.325***	−0.257**	−0.211**	−0.402***	0.258**	−0.235**
	(0.103)	(0.128)	(0.102)	(0.130)	(0.117)	(0.110)
Mean of Y	0.269	0.528	0.843	4.632	0.301	0.351
Observations	2184	2184	2184	1624	1802	1802

Notes: Each column and panel displays coefficients from separate OLS regressions. See Table 4 for a description of the dependent variables in each column. Each panel reports results using different samples to compute the in-hospital mortality rate. Panel A repeats the main specification in Panel B of Table 5, Panel B uses admission from all cohorts during the sample period to compute the mortality rates, and Panel C uses all admissions between 1870 and 1901. All regressions include household, birth year, and birth order fixed effects, as well as indicators for the younger sibling in a sibling pair and the census year in which the outcome variable is observed. Standard errors are clustered at the household level.

Table A11: Sensitivity to definition of in-hospital mortality rate

	Individual occupational success				Mobility	
	(1) White collar	(2) Skilled +	(3) Semi-skilled +	(4) Ln wage	(5) Class ↘	(6) Class ↗
Panel A: Minimum mortality rate across admission conditions						
Mortality Rate	−0.294**	−0.238	−0.171	−0.386***	0.226	−0.213*
	(0.124)	(0.161)	(0.132)	(0.146)	(0.156)	(0.126)
Panel B: Mortality rate weighted by frequency of each condition						
Mortality Rate	−0.785**	−0.660	−0.281	−1.055***	0.433	−0.445
	(0.345)	(0.444)	(0.326)	(0.349)	(0.417)	(0.322)
Panel C: Mortality rate of most frequently reported condition						
Mortality Rate	−0.303***	−0.208	−0.114	−0.350***	0.194	−0.136
	(0.115)	(0.152)	(0.119)	(0.134)	(0.142)	(0.122)
Mean of Y	0.269	0.528	0.843	4.632	0.301	0.351
Observations	2184	2184	2184	1624	1802	1802

Notes: Each column and panel displays coefficients from separate OLS regressions. See Table 4 for a description of the dependent variables in each column. Each panel reports results using different definitions to compute the in-hospital mortality rate for the admitted condition. Panel A uses the lowest mortality among the admitted conditions, Panel B weights the mortality rate of each condition by its frequency in the full sample of admissions, and Panel C uses the mortality rate of the most frequently reported condition. All regressions include household, birth year, and birth order fixed effects, as well as indicators for the younger sibling in a sibling pair and the census year in which the outcome variable is observed. Standard errors are clustered at the household level.

Table A12: Sensitivity to definition of treatment indicator

	Individual occupational success				Mobility	
	(1) White collar	(2) Skilled +	(3) Semi-skilled +	(4) Ln wage	(5) Class ↘	(6) Class ↗
Panel A: Restricting attention to first observed admission						
Mortality rate	−0.309***	−0.314**	−0.190*	−0.390***	0.266**	−0.173
	(0.107)	(0.128)	(0.104)	(0.123)	(0.119)	(0.112)
Mean of Y	0.268	0.528	0.843	4.633	0.301	0.349
Observations	2176	2176	2176	1618	1794	1794
Panel B: Specifying treatment as sum of all hospital admissions						
Mortality rate	−0.375***	−0.350**	−0.241**	−0.351***	0.273**	−0.222*
	(0.117)	(0.137)	(0.116)	(0.132)	(0.131)	(0.124)
Mean of Y	0.269	0.526	0.842	4.629	0.298	0.352
Observations	1920	1920	1920	1416	1596	1596

Notes: Each column and panel displays coefficients from separate OLS regressions. See Table 4 for a description of the dependent variables in each column. Panel A defines the in-hospital mortality treatment using only the first observed admission for each patient, and Panel B defines the treatment as the sum of all admissions. All regressions include household, birth year, and birth order fixed effects, as well as indicators for the younger sibling in a sibling pair and the census year in which the outcome variable is observed. Standard errors are clustered at the household level.

Table A13: Sensitivity to restrictions on county of enumeration in census

	Individual occupational success				Mobility	
	(1) White collar	(2) Skilled +	(3) Semi-skilled +	(4) Ln wage	(5) Class ↘	(6) Class ↗
Panel A: Restricting sample to County of London						
Mortality rate	−0.351***	−0.282**	−0.214**	−0.357***	0.312**	−0.269**
	(0.107)	(0.134)	(0.101)	(0.123)	(0.121)	(0.114)
Mean of Y	0.271	0.534	0.857	4.645	0.301	0.352
Observations	1920	1920	1920	1402	1618	1618
Panel B: Increasing sample to all counties of England						
Mortality rate	−0.320***	−0.260**	−0.178*	−0.373***	0.241**	−0.193*
	(0.102)	(0.128)	(0.100)	(0.116)	(0.119)	(0.109)
Mean of Y	0.265	0.523	0.836	4.623	0.306	0.347
Observations	2298	2298	2298	1716	1848	1848

Notes: Each column and panel displays coefficients from separate OLS regressions. See Table 4 for a description of the dependent variables in each column. Panel A excludes households residing outside the County of London from the main regression sample, while Panel B adds all households residing in England to the regression sample. All regressions include household, birth year, and birth order fixed effects, as well as indicators for the younger sibling in a sibling pair and the census year in which the outcome variable is observed. Standard errors are clustered at the household level.

Table A14: Sensitivity to matching on hints

	Individual occupational success				Mobility	
	(1) White collar	(2) Skilled +	(3) Semi-skilled +	(4) Ln wage	(5) Class ↘	(6) Class ↗
Panel A: Restricting to observations that match on birth parish						
Mortality rate	−0.356***	−0.261	−0.050	−0.321**	0.213	−0.298*
	(0.135)	(0.171)	(0.131)	(0.157)	(0.145)	(0.165)
Mean of Y	0.272	0.526	0.842	4.654	0.284	0.345
Observations	1260	1260	1260	968	1066	1066
Panel B: Restricting to observations that are matched via genealogy hints						
Mortality rate	−0.413***	−0.449***	−0.032	−0.443**	0.188	−0.235
	(0.143)	(0.169)	(0.130)	(0.190)	(0.168)	(0.161)
Mean of Y	0.287	0.555	0.848	4.650	0.278	0.340
Observations	1146	1146	1146	854	950	950

Notes: Each column and panel displays coefficients from separate OLS regressions. See Table 4 for a description of the dependent variables in each column. Panel A restricts attention to households with a patient and sibling that both have matching birth parishes in both census years, and Panel B restricts the sample to households with both members matched via genealogy hints. All regressions include household, birth year, and birth order fixed effects, as well as indicators for the younger sibling in a sibling pair and the census year in which the outcome variable is observed. Standard errors are clustered at the household level.

Table A15: Results by hospital. Sibling fixed effects.

	Individual occupational success				Mobility	
	(1) White collar	(2) Skilled +	(3) Semi-skilled +	(4) Ln wage	(5) Class ↘	(6) Class ↗
Panel A: Effects of admission by hospital						
Barts patient	−0.060***	−0.047*	−0.016	−0.032	0.018	−0.027
	(0.022)	(0.026)	(0.020)	(0.027)	(0.025)	(0.024)
GOSH patient	−0.015	−0.038	−0.025	−0.058*	0.021	0.008
	(0.028)	(0.032)	(0.023)	(0.034)	(0.028)	(0.030)
Panel B: Interacting mortality rate index with hospital indicators						
Barts × MR	−0.510***	−0.391**	−0.248*	−0.415***	0.424**	−0.422***
	(0.128)	(0.180)	(0.144)	(0.145)	(0.169)	(0.140)
GOSH × MR	−0.129	−0.110	−0.076	−0.351*	0.044	0.101
	(0.171)	(0.188)	(0.148)	(0.187)	(0.168)	(0.180)
Mean of Y	0.269	0.528	0.843	4.632	0.301	0.351
Observations	2184	2184	2184	1624	1802	1802

Notes: Each column and panel displays coefficients from separate OLS regressions. See Table 4 for a description of the dependent variables in each column. Panel A presents coefficients from a specification with separate indicators for admission at Barts or GOSH hospitals, and Panel B interacts these indicators with the in-hospital mortality rate of the admitted condition. All regressions include household, birth year, and birth order fixed effects, as well as indicators for the younger sibling in a sibling pair and the census year in which the outcome variable is observed. Standard errors are clustered at the household level.

Table A16: Effects of admission severity by age-at-admission group ([0,4] vs. [5,11])

	Individual occupational success				Mobility	
	(1) White collar	(2) Skilled +	(3) Semi-skilled +	(4) Ln wage	(5) Class ↘	(6) Class ↗
Panel A: Interacting mortality index with age-at-admission groups						
MR × [0-4]	−0.097	−0.038	−0.027	−0.127	0.174	−0.003
	(0.146)	(0.182)	(0.144)	(0.159)	(0.178)	(0.161)
MR × [5-11]	−0.561***	−0.467**	−0.299**	−0.566***	0.308*	−0.366**
	(0.148)	(0.181)	(0.144)	(0.180)	(0.162)	(0.150)
Panel B: Interacting acute indicator with age-at-admission groups						
Acute × [0,4]	−0.021	−0.081*	0.020	−0.062	0.078*	−0.040
	(0.042)	(0.046)	(0.041)	(0.047)	(0.044)	(0.044)
Acute × [5,11]	−0.147***	−0.099***	−0.050*	−0.104***	0.052	−0.054
	(0.033)	(0.038)	(0.030)	(0.037)	(0.037)	(0.034)
Non-acute × [0,4]	−0.019	0.005	−0.053*	−0.030	0.030	0.029
	(0.036)	(0.043)	(0.032)	(0.041)	(0.038)	(0.041)
Non-acute × [5,11]	0.022	0.005	0.000	0.032	−0.039	0.016
	(0.029)	(0.032)	(0.023)	(0.037)	(0.030)	(0.030)
Mean of Y	0.269	0.528	0.843	4.632	0.301	0.351
Observations	2184	2184	2184	1624	1802	1802

Notes: Each column and panel displays coefficients from separate OLS regressions. See Table 4 for a description of the dependent variables in each column. Panel A presents coefficients from a specification with separate indicators for admission at ages 0 to 4 and 5 to 11 interacted with the in-hospital mortality rate of the admitted condition, and Panel B interacts the age-at-admission indicators with indicators for acute and non-acute conditions. All regressions include household, birth year, and birth order fixed effects, as well as indicators for the younger sibling in a sibling pair and the census year in which the outcome variable is observed. Standard errors are clustered at the household level.

Data Appendix

Constructing in-hospital mortality rates

We develop a measure of severity for each hospital admission based on the in-hospital mortality rate for the admitted conditions. The admission registers contain a description of the conditions from which patients suffered when admitted to the hospital. The raw text strings are cleaned and components are assigned to one of eight categories:

1. Diseases (e.g. bronchitis, typhoid fever)
2. Symptoms (e.g. obstruction, pain, swelling)
3. Body parts (e.g. joint, knee, toe)
4. Surgical procedures (e.g. amputation, excision, operation)
5. Accidents or injuries (e.g. burn, fracture, run over, shot)
6. Objects (e.g. poisoning, swallowed a spoon)
7. Location (e.g. left, middle, under)
8. Severity (e.g. acute, chronic, traumatic)

Among the 88 percent of patients for whom a condition can be identified, 46 percent of cause-of-admission descriptions contain multiple components. The admission registers also record whether a patient dies in the hospital, either in a "Notes" field for St. Bartholomew's Hospital, or as an outcome in the "Result" field for Great Ormond Street and Guy's Hospitals. We compute an in-hospital mortality rate for each unique string component, treating each component for a given admission as a separate observation. Mortality rates are computed for three separate samples:

1. Cohorts in sample (Patients from the 1874 to 1890 birth cohorts)

2. Sample years (Patients admitted between 1874 and 1901)

3. All admissions

In all cases, we restrict attention to male (and ambiguous gender) patients and calculate hospital-specific mortality rates. When a cause-of-admission description contains multiple components, we use four rules to assign a mortality rate for that particular admission:

1. Maximum mortality rate across components

2. Minimum mortality rate across components

3. Mortality rate for the most frequently occurring condition (within the same hospital and sample)

4. Frequency weighted mortality rate

In our main specifications we use the cohorts in our sample to calculate mortality rates and assign the maximum mortality rate across multiple reported components. Table A8 provides an example in which mortality rates are computed for the cause of admission string "Abd. pain, ?Enteric fever". Separate mortality rates are computed for "abdomen", "pain", and "enteric fever". In the main specification, we assign a mortality rate of 0.148 for this admission, based on the fraction of admissions containing the word "abdomen" (or "abdominal") among patients in our sample cohorts.

Method of constructing unique patient identifers

In contrast to modern day hospital record keeping, the nineteenth-century hospital admission records did not contain unique patient identifers. We proceed in two stages to construct a proxy for a patient identifier. First, we take unique combinations of first name, surname, birth year, and district of residence. Given that the hospital records only report age in years on the day of admission, we infer the birth year from the midpoint of the bounds on the birth date after subtracting the reported age from the year of admission. After matching the hospital admission records to the census, we group together admissions matched to the same individual in the census and use these new groupings of admission as patient identifiers.

Data Linkage

The process of constructing the linked data set begins by locating the childhood household of residence for the hospital patients in the 1871 through 1901 decennial censuses of England. We match records using first name, middle initial (when available), surname, and age in years on the day of census enumeration. We search for the hospital patients in the census closest to the date of admission in which the patient would have been alive on the day of enumeration.[63]

When linking to the 1891 and 1901 censuses, we search for the records in the databases of transcribed census records on a genealogy website (Ancestry.com, 2005a,b, 2011). A different procedure is used in matching the hospital records to the 1881 census of England, in which case the North Atlantic Population Project (NAPP) have publicly released the complete count file (Schurer and Woollard, 2003; Minnesota Population Center, 2008). The first stage of the linking procedure for the 1881 census involves taking as a potential match any pairs of observations that meet at least one of the following criteria:

(i) First letter of given name matches AND surname SOUNDEX code matches AND age in years is within 2 years.

[63]Hospital patients who report an age in years on admission that corresponds to a birth less than two years following a given decennial census are also searched for in that census to account for possible recall or transcription errors in the age variable.

(ii) First letter of surname matches AND given name SOUNDEX code matches AND age in years is within 2 years.

Steps (i) and (ii) are repeated by replacing the SOUNDEX code with original name string standardized by the Atack-Bateman implementation of the NYSIIS algorithm, or with the first three consonants of the original name string. The union of observations across the criteria are taken as potential matches to the 1881 census.

The second stage of the matching algorithm, which is applied separately for each linkage between the hospital records and a decennial census, consists of the following steps:

1. Exclude observations that have different non-missing middle initials, and rank the remaining observations based on match quality.

2. Require the Levenshtein distances of the first name and surname weighted by the length of each string to be less than 0.1, and the reported ages to be within one year across sources.

3. Prioritize observations with a closer match on age, then accept observations that uniquely minimize the length-weighted Levenshtein distance averaged over the first name and surname.

4. If a hospital record matches to multiple records in the census, exclude observations that do not match on district of residence.

5. Drop any remaining hospital records that do not have a unique match within a given census.

The matching procedure is complicated by the fact that none of the sources report an individual's year of birth. Instead, the hospital admission records report the age in years of each patient on the day of admission, while the census records report the age in years on the most recent birthday, as of the date of enumeration. We compute bounds on the date of birth under the assumption that the patient's most recent birthday was on the day of

admission, or was up to a year prior. Then, we compute bounds on the age in years on the day of census enumeration, and the minimum absolute deviation with respect to the age in years reported in the census for each potential match.

The search for an individual's census record during adulthood cannot be confined to a specific geographic location if one wishes to observe individuals to migrate. Thus, an individual's place of birth is also used to match childhood census records to adulthood. Although census enumerators were tasked to record an individual's parish and county of birth, they often recorded location names that did not correspond to parishes. Moreover, many birth place entries in 1911 have been transcribed as "London, United Kingdom," even when information on the parish of birth is recorded in the original document. Given these restrictions, we require entries to match on county of birth, but only exclude cases with non-matching parish of birth when the parish can be identified in both sources.