Lifetime and intergenerational consequences of poor childhood health

Krzysztof Karbownik and Anthony Wray

Online Appendix

For Online Publication: Online Appendix

This document contains supplementary materials for the paper "Lifetime and intergenerational consequences of poor childhood health" by Krzysztof Karbownik and Anthony Wray published in the *Journal of Human Resources*.

A Data linking procedure

In this section we describe the linking algorithms used to match inpatient admission records to census records and to match individuals across censuses. We explain the census-to-census linkage in detail as it is the most basic procedure and it is the basis for the hospital-to-census linkage with some modifications. The description applies to the linkage of male patients and their siblings for the analysis of long-run occupational outcomes. Where applicable, we indicate how the linkage procedure differs for the analysis of mechanisms (participation in schooling and disability) for which only a hospital-to-census linkage is required. We describe the procedure for linking men and women to study effects on the probability of being enumerated as a single adult in Online Appendix Section A.5, and the procedure for linking fathers in Online Appendix Section A.8.

A.1 Census-to-census linkages

We use the complete count data for the Censuses of England from the I-CeM project to create five linked samples: 1881 to 1901, 1881 to 1911, 1891 to 1901, 1891 to 1911, and 1901 to 1911 (Schurer and Higgs 2020, 2022).[1] In each case, we start with all males aged 0 to 21 in the base year census. Census-to-census linkages are based on time-invariant characteristics such as first name, surname, birth year, and county of birth.[2] We begin by separating given names into first and middle names, and then standardize diminutives and common nicknames of first names to their proper equivalents. We follow the procedure in Parman (2015) and construct the Phonex codes for the first and last names in each data set, which enables us to allow for differences in the spelling of phonetically similar names across data sets that might arise from factors such as typographical errors.[3] Prior to the implementation of the matching algorithm, we perform a "blocking" step in which the two data sets are joined using four blocking variables: the Phonex

[1]We also link the 1881 and 1891 censuses but only use this sample in the auxiliary analysis of fathers described in Online Appendix Section A.8.

[2]We choose not to match on birth parish during the initial step given that the variable is a non-standardized text string and parish boundaries changed significantly over the time period of study (Schürer and Day 2019).

[3]See Dahis, Nix and Qian (2020) and the supplemental materials to Parman (2015) for discussions of the Phonex algorithm.

code of the first and last names, age in years when enumerated in the later census, and county of birth (Christen 2012).

The linkage procedure draws on elements of the methods pioneered by Ferrie (1996) and utilized by Abramitzky, Boustan and Eriksson (2012), modifications developed by Feigenbaum (2016) and Mill and Stein (2016), and recommendations made by Bailey et al. (2020).[4] It proceeds as follows:

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

2. Drop all pairs of linked observations that do not have matching Phonex codes or county of birth, while allowing for discrepancies in the reported age of up to 3 years.

3. Compute the Jaro-Winkler score between the first names and last names in each pair of observations. Discard all pairs with a Jaro-Winkler score less than 0.75 for either the first or last name.[5] For first names, we use both the original string and the edited string, and take the highest Jaro-Winkler score among all possible comparisons of first name strings between the two censuses.

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

5. Prioritize linked observations that match on birth parish.

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

[4] While the linking methods used in this paper are not exact replications of traditional methods, the approach of incorporating features from different methods is validated by the findings of Bailey et al. (2020) that using a combination of samples generated with the Ferrie (1996) and Feigenbaum (2016) methods results in a much lower Type I error rate.

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

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

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

We present linkage rates separately for all census-to-census linkages in Online Appendix Table A22. The share of unique matches ranges from 51 to 64 percent across the set of census pairs, with higher match rates for censuses that are closer together, especially those that are only 10 years apart. The census-to-census linkage rates typically found in the literature using complete-count U.S. census data are somewhat lower. This difference can be explained in part by applications with longer windows of time between censuses, typically 30 to 40 years, where sample attrition is of greater concern. Furthermore, the U.S. censuses have less precise information on birth place, at the state level instead of county or parish, which reduces the likelihood of finding unique matches.

A.2 Hospital-to-census linkage

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

We link each hospital admission record to the 1881, 1891, and 1901 censuses provided that the admission occurred within 10 years of the census enumeration date. We do so for both boys and girls. We use information on the age in years on the day of the hospital admission to determine the age in years on the days of census enumeration. As with the census-to-census linkages, we require the age to differ by no more than 3 years between sources. In the absence of information on place of birth, we prioritize linkages of records that match on county of residence, but we do not require either district or county of residence to match since individuals moved often, even in short time windows between hospital admission and census enumeration. We discuss the overall linkage rates from hospital records to censuses in Section III.B.

A.3 Multiple linkages to a sample of unique individuals

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

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

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

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

Among those patients linked to census records during childhood, we update the unique identifiers based on the census linkages. In a small number of cases, admissions of patients with different proxy identifiers are linked to the same census record in either 1881, 1891, or 1901, and we consider them to be the same individual. When we conduct the second linkage to census records during adulthood, we further consolidate the proxy identifiers. For example, if one admission record is linked to the 1881 census, and another record is linked to the 1891 census, and both census records are linked to the same individual in either the 1901 or 1911 census, then we consider the two admission records to belong to the same patient. As illustrated in Table 1, many patients are linked to more than one census, with hospital-to-census linkage rates ranging from 25 to 28 percent for each of the 1881 through 1901 censuses, and 34 percent of patients matched to any census.

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

1. Choose the census closest to the admission year.

2. Select the census record with the smallest deviation in age between the hospital admission record and the childhood census.

3. Choose the childhood census record linked to the latest census year during adulthood (1901 or 1911).

4. Choose the earliest childhood census record (1881, 1891 or 1901).

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

A.4 Patient-sibling comparisons

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

1. Drop households with multiple patients.

2. Among successfully matched male siblings, keep the sibling who is closest in age to the hospital patient. In the cases where we link both an older and younger sibling with the same age gap in comparison to the hospital patient, we choose the older sibling if the patient's unique identifier in the I-CeM complete count files is an even number and the younger sibling if it is an odd number, in order to avoid biasing the sibling fixed effects comparisons to either younger or older siblings.

A.5 Linking to long-run marital status: Likelihood of enumeration as a single adult

Here we describe the process of generating the samples used to study the effects on the likelihood of being enumerated as a single adult for men and women that are presented in Table 3. The starting point for the analysis of women in column 1 is the sample of girls admitted to a hospital as children and linked to 1881, 1891, or 1901 Censuses of England, along with their sisters who were enumerated in the same household, which we discussed above in Online Appendix Section A.2. In comparison to the main analysis, we impose a further restriction that individuals must be no older than age 15 when enumerated in the childhood census. Online Appendix Figure A15 shows that only a trivial number of girls below the age of 16 are married, based on the

1911 complete-count census data, and thus this restriction ensures that our sample of patients and their siblings is not biased by selection into marriage. As with the main analysis we match patients to multiple censuses and select a single linked pair of records for each patient based on the following steps:

1. Choose the census closest to the admission year.
2. Select the census record with the smallest deviation in age between the hospital admission record and the childhood census.
3. Choose the earliest census data (1881 or 1891).

We restrict the sample to siblings within 5 years of age of the patient. We then attempt to link patients and siblings to censuses 10, 20, or 30 years later, restricting to women who are single and at least 18 years old in the later census. We ensure that linked women are in fact single by excluding individuals who report ever being married. The outcome variable in the regressions is an indicator for a successful linkage to a single women in at least one of the later censuses. The sample of men in column 2 is constructed in identical fashion. In column 3 we restrict the sample to men who are successfully linked to one of the later censuses irrespective of their marital status.

A.6 Linked samples for analysis of participation in schooling

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

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

When constructing the linked sample, we match both male and female patients to either the 1881 or the 1891 censuses if they were admitted to the hospital prior to enumeration and were ages 5 to 10 years old at the time of the census. This implies that the estimation sample includes individuals from the 1871 to 1876 birth cohorts linked to the 1881 census and individuals from the 1881 to 1886 birth cohorts linked to the 1891 census. This procedure ensures that we choose the highest quality match for the analysis sample, before we impose additional restrictions so that we observe the individual in the census during the compulsory schooling years and after the hospital admission.

A.7 Linked samples for analysis of disability

For the childhood disability analysis, our sample consists of male and female patients linked from the hospital records to one of the 1881 through 1911 censuses up to 10 years post admission, as well as their siblings located in the same household in the census. In the same manner as in the main occupational analysis, we restrict the sample to patients born between 1870 and 1890 and admitted at ages 0 to 11. We further require that both patients and siblings be 21 years or younger at the time of census enumeration.

When we turn to disability in adulthood we can only include men in the sample due to name changes at marriage by women. The construction of the long-run sample follows the procedure outlined for the long-run occupational outcomes, but we additionally include individuals with missing occupational outcomes. In a separate exercise, we examine whether hospitalization is associated with within-household differences in the likelihood of disability prior to admission. In this case, the sample construction is identical to the childhood disability sample, with the exception that patients are linked from the hospital records to a census up to 10 years before admission.

A.8 Linked sample of fathers

Thus far, we have described methods for linking population censuses over time and to hospital admission records, which we use to generate our main analysis samples. In an auxiliary analysis discussed in Section V.A and reported in Online Appendix Figure A7, we also make use of a sample of fathers linked between adjacent censuses. The starting points for the linked sample of fathers are the samples of children age 0 to 21 linked from the 1881 to 1891 and 1891 to 1901 censuses. We append to these samples pairs of census records that are linked to the same hospital patients. We require children to be enumerated in the same household as their father and for their father to report an occupation in both censuses.

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

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

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

B Transition matrices and the interpretation of absolute and relative mobility measures

B.1 Occupational transition matrix

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

B.2 Interpretation of absolute and relative mobility measures

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

If we consider the first household, the relative outcome of upward mobility will differ between patient and sibling, since only the sibling moves up in status relative to his father, whereas in the second household there is no difference in relative mobility between patient and sibling, since both improve in occupational rank compared to their father. Importantly, the coefficients on upward and downward mobility are not necessarily of equal magnitude and of opposite sign. Households in which one sibling moves up in status and the other moves down relative to the father will have differences between patient and sibling in both the upward and downward mobility outcomes. Yet, a household in which one sibling moves up and and the other stays in the same occupation as their father will have positive upward but zero downward mobility, while a household with one sibling moving down and the other staying in the same occupation as their father will have positive downward but zero upward mobility. The latter case highlights a limitation of the relative mobility outcome in that it does not distinguish between differences in the degree of upward mobility, i.e. rising by two social ranks is better than rising by one rank. Nonetheless, we include the relative mobility measures since the interpretation of these outcomes is comparable to the analysis of transition matrices that has been explored extensively in the occupational mobility literature (Long and Ferrie 2013; Feigenbaum 2018; Pérez 2019).

Regardless, we also examine absolute mobility outcomes such as an indicator for white collar occupational status. For the second household in the example, there is a difference in absolute outcomes between patient and sibling, given that ending up in the middle of the occupational distribution is worse than rising to the top. However, in the first household, there is no difference in white collar status between patient and sibling, as neither attains white collar status. In the final empirical sample, 25 to 28 percent of households have variation in the mobility measures while 20 to 39 percent have variation in the own rank outcomes. An indicator for households with patient-sibling pairs that have variation in outcomes is uncorrelated with potentially confounding explanatory variables such as father's age categories, sibship size, characteristics of the patient's linkage to the census, and whether the household is located in the hospital's catchment area. As the example illustrates, in specifications with a relative mobility outcome, a different set of households contribute non-zero differences to the sibling fixed effects estimator compared to specifications with the absolute occupation outcomes. Thus, the estimates will differ and capture complementary but distinct measures of occupational success.

C Health Deficiency Index

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

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

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

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

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

$$P(\textit{Death in hospital})_{nhay}^{g} = \alpha + \theta_h + \delta_a + \gamma_y + \epsilon_{nhay}^{g} \tag{1}$$

where n^g indexes individual in-patient admissions for gender g, h indexes hospitals, a indexes age in years at admissions, and y indexes the year of admission. The dependent variable is an indicator that takes the value of one when a patient dies in the hospital. We include hospital (θ), age at admission (δ), and year of admission (γ) fixed effects. We save the residuals $\widehat{\epsilon_{nhay}^{g}}$ from the regression to use as an input in the next step of computing the health deficiency index.

The estimation excludes observations with no diagnosis and diagnoses with at least 25 observations for which there is no variation in observed inpatient mortality.[6] Next, we assign patients the average residual mortality risk for their primary diagnosis as a proxy for childhood health. For each diagnosis d_j^g of gender g, we compute the following:

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

which is the average unexplained mortality risk across all admissions of gender g with primary diagnosis equal to d_j^g. Finally, we compute the health deficiency index by the following steps:

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

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

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

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

D Weighting

Here, we describe the procedure used to re-weight the data by observable characteristics of all patients in the hospital records. The results are reported in Online Appendix Tables A10 and A11. Given that the hospital records represent the starting point for our sample construction, we take inpatients at risk of being linked to the census as the baseline population. The at-risk population consists of patients born between 1870 and 1890 and admitted to one of the three hospitals in our sample at ages 0 to 11 between 1870 and 1902.

We re-weight the data to ensure that the final empirical samples match the proportions in the hospital records based on the following observable characteristics: age at admission, year at admission and place of residence. In specifications involving male and female patients we also re-weight by gender. Our procedure follows Abramitzky et al. (2021) and Black et al. (2021) in that we compute quintile bins for the continuous variables: age at admission (0-1, 2-3, 4-5, 6-8, 9-11) and year of admission (1870-82, 1883-88, 1889-93, 1894-98, 1899-02). Likewise, residential location is measured by the place of residence of a patient at the time of admission. We include indicators for registration districts of London inside and outside a hospital's catchment area, and counties of Greater London (Essex, Kent, Middlesex, and Surrey), with remaining counties as the excluded category. We estimate a probit regression on the population of hospital admissions with the dependent variable equal to one if an individual appears in the final empirical sample. Weights are computed as:

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

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

References

Abramitzky, Ran, Leah Platt Boustan and Katherine Eriksson. 2012. "Europe's Tired, Poor, Huddled Masses: Self-Selection and Economic Outcomes in the Age of Mass Migration." *American Economic Review* 102(5):1832–1856.

Abramitzky, Ran, Leah Platt Boustan, Katherine Eriksson, James J. Feigenbaum and Santiago Pérez. 2021. "Automated Linking of Historical Data." *Journal of Economic Literature* 59(3):865–918.

Bailey, Martha J., Connor Cole, Morgan Henderson and Catherine Massey. 2020. "How Well Do Automated Linking Methods Perform? Lessons from US Historical Data." *Journal of Economic Literature* 58(4):997–1044.

Black, Sandra, Sanni Breining, David Figlio, Jonathan Guryan, Krzysztof Karbownik, Helena Skyt Nielsen, Jeffrey Roth and Marianne Simonsen. 2021. "Sibling Spillovers." *Economic Journal* 131(633):101–128.

Chatto and Windus. 1897. "Sixty Years of Healing." *Belgravia: A London Magazine* 93:203–233.

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

Dahis, Ricardo, Emily Nix and Nancy Qian. 2020. "Choosing Racial Identity in the United States, 1880-1940." NBER Working Paper 26465.

Feigenbaum, James. 2018. "Multiple Measures of Historical Intergenerational Mobility: Iowa 1915 to 1940." *Economic Journal* 128(612):F446–F481.

Feigenbaum, James J. 2016. "Automated Census Record Linking: A Machine Learning Approach." Working Paper.

Ferrie, Joseph P. 1996. "A New Sample of Males Linked from the Public Use Micro Sample of the 1850 U.S. Federal Census of Population to the 1860 U.S. Federal Census Manuscript Schedules." *Historical Methods* 29(4):141–156.

Long, Jason and Joseph Ferrie. 2013. "Intergenerational Occupational Mobility in Great Britain and the United States since 1850." *American Economic Review* 103(4):1109–1137.

Mill, Roy and Luke C. D. Stein. 2016. "Race, Skin Color, and Economic Outcomes in Early Twentieth-Century America." SSRN Working Paper 2741797.

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

Pérez, Santiago. 2019. "Intergenerational Occupational Mobility across Three Continents." *Journal of Economic History* 79(2):383–416.

Schürer, K and Joe Day. 2019. "Migration to London and the Development of the North–South Divide, 1851–1911." *Social History* 44(1):26–56.

Schurer, Kevin and Edward Higgs. 2020. *Integrated Census Microdata (I-CeM), 1851-1911. [data collection]*. UK Data Service. SN: 7481. http://doi.org/10.5255/UKDA-SN-7481-2.

Schurer, Kevin and Edward Higgs. 2022. *Integrated Census Microdata (I-CeM) Names and Addresses, 1851-1911: Special Licence Access. [data collection]. 2nd Edition.* UK Data Service. SN: 7856. http://doi.org/10.5255/UKDA-SN-7856-2.

E Appendix Tables

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

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

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

Source: Chatto and Windus (1897).

Table A2: Common causes of admission in hospital population by gender and in long-run occupational sample

	Hospital male population				Hospital female population				LR sample (Males)	
Cause of admission	Frequency	Percent	Mortality rate	Cause of admission	Frequency	Percent	Mortality rate	Cause of admission	Frequency	Percent
Abscess	1,901	4.52	0.04	Chorea	1,474	4.99	0.01	Abscess	105	4.89
Pneumonia	1,513	3.59	0.11	Abscess	1,297	4.39	0.05	Pneumonia	65	3.03
Diphtheria	1,499	3.56	0.36	Diphtheria	1,252	4.24	0.39	Fracture	63	2.94
Tubercular Disease	1,422	3.38	0.04	Tubercular Disease	972	3.29	0.04	Bronchitis	57	2.66
Bronchopneumonia	1,120	2.66	0.32	Pneumonia	838	2.84	0.12	Phimosis	54	2.52
Bronchitis	999	2.37	0.17	Bronchopneumonia	801	2.71	0.28	Diphtheria	45	2.10
Fracture	939	2.23	0.03	Bronchitis	751	2.54	0.16	Chorea	44	2.05
Meningitis	718	1.71	0.77	Cleft Palate	558	1.89	0.00	Typhoid Fever	42	1.96
Empyema	686	1.63	0.14	Meningitis	505	1.71	0.77	Tubercular Disease	40	1.86
Chorea	618	1.47	0.01	Morbus Cordis	498	1.69	0.18	Injury	38	1.77
Fever	613	1.46	0.11	Tuberculosis	439	1.49	0.57	Empyema	37	1.72
Phimosis	601	1.43	0.01	Typhoid Fever	426	1.44	0.06	Cleft Palate	35	1.63
Injury	581	1.38	0.04	Phthisis	401	1.36	0.22	Rheumatism	35	1.63
Typhoid Fever	569	1.35	0.07	Empyema	395	1.34	0.12	Talipes	35	1.63
Harelip	556	1.32	0.02	Rheumatism	392	1.33	0.02	Rickets	32	1.49
Tuberculosis	539	1.28	0.57	Fever	383	1.30	0.10	Necrosis	31	1.44
Cleft Palate	535	1.27	0.01	Rickets	353	1.19	0.03	Harelip	29	1.35
Rheumatism	524	1.24	0.02	Burn	322	1.09	0.30	Fever	28	1.30
Talipes	517	1.23	0.00	Disease Hip	317	1.07	0.02	Morbus Cordis	28	1.30
Morbus Cordis	456	1.08	0.16	Fracture	312	1.06	0.01	Scarlet Fever	25	1.16
Burn	451	1.07	0.25	Caries Spine	308	1.04	0.04	Disease Knee	24	1.12
Rickets	440	1.05	0.05	Morbus Coxae	306	1.04	0.02	Pleurisy	24	1.12
Diarrhea	418	0.99	0.24	Harelip	296	1.00	0.01	Nephritis	23	1.07
Laryngitis	412	0.98	0.16	Diarrhea	288	0.97	0.29	Caries Spine	21	0.98
Necrosis	409	0.97	0.03	Talipes	282	0.95	0.01	Diarrhea	21	0.98
Total (top 25)	19,036	45.22	0.15	Total (top 25)	14,166	47.96	0.16	Total (top 25)	981	45.70
Outside top 25	23,064	54.78	0.11	Outside top 25	15,387	52.04	0.11	Outside top 25	1,165	54.30

Notes: This table lists the 25 most common causes of admission in the populations of hospitalized male and female patients, as well as the sample of males used in the long-run occupational analysis. The hospital populations consist of all admissions by patients born between 1870 and 1902 and admitted at ages 0 to 11 from 1870 to 1902 at GOSH, Barts, or Guy's Hospitals. The causes of admission are tabulated after cleaning the text strings transcribed from the admissions registers. The mortality rate refers to the share of admissions in which a patient died in the hospital. The long-run (LR) sample of males refers to the set of 2,146 hospital admissions reported in column 4 of Table 1, which correspond to the 1,849 male patients included in the main analysis in Table 2. The mortality rate is not shown for the long-run sample since it only includes patients who survived until adulthood. In the LR sample, admissions for hip disease and eczema also occurred 21 times.

16

Table A3: Common occupational titles by occupational class

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

Notes: This table lists the five most common occupations in each of four occupational classes for the final sample of patients and siblings used in the main analysis of long-run occupational outcomes reported in Table 2. Column 1 combines professional, managerial and clerical occupations, which correspond to classes 1 to 5 in the Historical International Social Class Scheme (HISCLASS), into a white collar class. Column 2 subsumes farmers into skilled workers (HISCLASS 6 to 8), column 3 displays semi-skilled workers (HISCLASS 9), and column 4 combines unskilled workers as well as low and unskilled farm workers (HISCLASS 10 to 12). See Section III.C for further details on the occupational classification.

Table A4: Intergenerational occupational elasticities

	Son's occupational status		
	(1)	(2)	(3)
	White collar	Skilled +	Log wage
Father's status	0.250***	0.219***	0.255***
	(0.001)	(0.001)	(0.001)
Mean of Y	0.240	0.488	4.589
N	2,053,932	2,053,932	2,055,293

Notes: This table presents estimates using data on males aged 0 to 11 linked from the 1881 to the 1911 complete-count census. In each column, the dependent variable is an indicator for the son's occupational status or the son's log occupational wage which are identical to the dependent variables in columns 3, 4, and 6 of Table 2. The treatment variable "father's status" varies across the columns and in each case is defined in an equivalent way as the dependent variable but for the father rather than the son. The regressions also control for an indicator for above-median sibship size, match quality dummies, as well as own and father's birth year fixed effects.

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

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

	Father's occupational class				Total	N
	White collar	Skilled	Semi-skilled	Unskilled		
			Panel A: Patients			
White collar	45.8	21.4	23.4	16.2	25.7	475
Skilled	21.1	38.7	24.1	22.8	27.7	512
Semi-skilled	22.2	26.9	38.2	31.9	30.4	563
Unskilled	11.0	13.0	14.3	29.1	16.2	299
N	356	561	568	364	1,849	
			Panel B: Siblings			
White collar	49.7	23.4	28.9	16.8	28.8	533
Skilled	18.3	38.9	21.7	19.0	25.7	475
Semi-skilled	23.0	28.2	36.6	34.6	31.0	574
Unskilled	9.0	9.6	12.9	29.7	14.4	267
N	356	561	568	364	1,849	
			Panel C: Patients and siblings			
White collar	47.8	22.4	26.1	16.5	27.3	1,008
Skilled	19.7	38.8	22.9	20.9	26.7	987
Semi-skilled	22.6	27.5	37.4	33.2	30.7	1,137
Unskilled	10.0	11.3	13.6	29.4	15.3	566
N	712	1,122	1,136	728	3,698	
			Panel D: Population			
White collar	46.8	20.1	17.6	14.2	22.7	486,600
Skilled	20.8	40.6	19.9	19.6	26.0	555,835
Semi-skilled	22.9	26.2	50.1	31.5	34.5	738,006
Unskilled	9.5	13.1	12.4	34.7	16.8	359,830
N	376,328	617,033	693,379	453,531	2,140,271	

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

Table A6: Effects on long-run social outcomes

	(1) Share unskilled in neighborhood	(2) Living with a parent	(3) Moved to new county	(4) Has any children	(5) Child participating in schooling
	Panel A: Effects of hospital admission				
Patient	0.002**	0.020	−0.011	−0.009	−0.001
	(0.001)	(0.013)	(0.013)	(0.014)	(0.035)
% effect	2.5	5.7	3.5	2.7	0.3
	Panel B: Effects of health deficiency index				
Health deficiency index	0.009**	0.043	−0.026	0.007	−0.067
	(0.003)	(0.043)	(0.043)	(0.045)	(0.114)
% effect (σ)	1.2	1.6	1.1	0.3	2.7
Mean of Y	0.095	0.352	0.314	0.354	0.320
N	4,312	4,312	4,312	4,312	832

Notes: This table presents sibling fixed effects estimates of the patient indicator (panel A) and the health deficiency index (panel B) on social outcomes observed at the time of census enumeration during adulthood. In column 1, the dependent variable is the share of individuals who work in unskilled occupations in the parish of residence during adulthood. Unskilled occupations are defined as HISCLASS ranks 10 to 12 and correspond to the outcome in column 5 of Table 2. In column 2, the dependent variable is an indicator for living in the same household as at least one parent. In column 3, it is an indicator for residing in a different county than when enumerated in the census during childhood. In column 4, it is an indicator for having any children. In column 5, it is an indicator for the child participating in schooling (see Section III.F for details on how the variable is coded). In columns 1 to 4, the sample first restricts attention to individuals in the main empirical sample in Table 2, with further restriction to households in which the outcome variable is not missing for both patient and sibling, but allowing sample individuals and their fathers to have missing occupations. Column 5 also requires both patient and sibling to have a child. See Table 2 for a description of the control variables. Standard errors are clustered by childhood household. Percent effects in panel B are computed as $\beta \times \sigma_{HDI}$ scaled by the dependent variable mean, where $\sigma_{HDI} = 0.13$ is the standard deviation of the health deficiency index in the hospital population.

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

Table A7: Heterogeneity by severity and age at admission

	(1) Class $\nearrow$	(2) Class $\searrow$	(3) White collar	(4) Skilled +	(5) Unskilled	(6) Log wage
Panel A: Interaction with above vs. below median HDI						
Patient $\times$ low-HDI	-0.031^{*}	0.027	-0.042^{**}	-0.030	0.028^{*}	-0.036^{**}
	(0.017)	(0.016)	(0.018)	(0.020)	(0.015)	(0.016)
Patient $\times$ high-HDI	-0.030	0.029	-0.045^{**}	-0.045^{*}	0.039^{**}	-0.029^{*}
	(0.020)	(0.018)	(0.020)	(0.024)	(0.017)	(0.017)
P-value	0.986	0.921	0.917	0.608	0.608	0.744
Panel B: Interaction with early (0-4) vs. late (5-11) childhood admission						
Patient $\times$ [0-4]	-0.059^{***}	0.022	-0.056^{**}	-0.057^{**}	0.041^{**}	-0.051^{***}
	(0.021)	(0.020)	(0.023)	(0.026)	(0.018)	(0.019)
Patient $\times$ [5-11]	-0.020	0.029^{*}	-0.041^{**}	-0.025	0.029^{**}	-0.024^{*}
	(0.017)	(0.016)	(0.017)	(0.020)	(0.014)	(0.015)
P-value	0.134	0.801	0.584	0.303	0.579	0.245
Mean of Y	0.357	0.260	0.273	0.539	0.153	4.629
N	3,698	3,698	3,698	3,698	3,698	3,698

Notes: This table presents sibling fixed effects estimates. Panel A displays estimates in which we interact the indicator for hospital patient with indicators for being admitted for conditions with above and below median values of the health deficiency index. Panel B interacts the indicator variable for hospitalization with separate indicators for early- (age 0 to 4) and late-childhood (age 5 to 11) admission, which are coded based on a patient's first observed admission to a hospital. The dependent variables in columns 1 to 6 correspond to those shown in Table 2. See Table 2 for a description of the control variables. Standard errors are clustered by childhood household.

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

Table A8: Descriptive statistics for hospital catchment areas

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

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

Table A9: Selection into hospitalization for male patients

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

Notes: This table presents OLS estimates using a sample that consists of individuals who were ages 0 to 5 and residing in the County of London when enumerated in the 1881, 1891 or 1901 censuses. The dependent variable is an indicator for a unique match to an inpatient hospital admission that occurred up to 10 years after the census enumeration date and when the individual was age 0 to 11 at the time of admission. The regressions also include age-at-enumeration by census-year fixed effects for patients and their fathers.
Point estimates marked ***, **, and * are statistically significant at the 1, 5, and 10 percent levels, respectively.

Table A10: Weighting by characteristics of patient population – Occupational outcomes and mechanisms

	Occupational						Schooling	Disability		
	(1) Class ↗	(2) Class ↘	(3) White collar	(4) Skilled +	(5) Unskilled	(6) Log wage	(7) Participation	(8) Pre-existing	(9) Childhood	(10) Long-run
	Panel A: Effects of hospital admission									
Patient	-0.030^{**}	0.025^{*}	-0.043^{***}	-0.036^{**}	0.027^{**}	-0.035^{***}	-0.023^{***}	0.090	0.772^{***}	0.810^{***}
	(0.014)	(0.013)	(0.015)	(0.017)	(0.012)	(0.012)	(0.008)	(0.116)	(0.190)	(0.244)
% effect	8.5	9.7	15.7	6.8	17.6	3.4	3.1	82.3	149.3	136.7
	Panel B: Effects of health deficiency index									
Health deficiency index	-0.104^{**}	0.073^{*}	-0.141^{***}	-0.115^{**}	0.089^{**}	-0.112^{***}	-0.050	0.065	2.385^{***}	2.013^{**}
	(0.047)	(0.045)	(0.048)	(0.056)	(0.039)	(0.041)	(0.034)	(0.268)	(0.604)	(0.848)
% effect (σ)	3.8	3.6	6.7	2.8	7.4	1.4	0.9	7.6	59.5	43.9
Mean of Y	0.355	0.260	0.273	0.538	0.155	4.630	0.729	0.109	0.517	0.592
N	3,698	3,698	3,698	3,698	3,698	3,698	3,040	9,096	11,312	4,312

Notes: This table displays sibling fixed effects estimates from regressions that re-weight the data by the following observable characteristics in the hospital records: admission age, admission year, and location of residence in all columns, as well as gender in columns 7 to 9. Re-weighting the data ensures that each linked sample matches the proportions in the population of inpatient hospital admissions. The weighting procedure is described in more detail in Online Appendix D. In panel A, the explanatory variable of interest is an indicator for hospital admission, and in panel B, it is the health deficiency index. Columns 1 to 6 present estimates for the outcomes variables shown in Table 2. Columns 7 to 9 display estimates for the schooling and disability mechanisms from the pooled gender samples in Table 4, while column 10 reports estimates for long-run disability using the linked sample of boys only. See Tables 2 and 4 for a description of the control variables. Standard errors are clustered by childhood household. Percent effects for log wages (column 6) are calculated using the formula $100 \times \exp(\beta) - 1$. Percent effects in panel B are computed as $\beta \times \sigma_{\text{HDI}}$ scaled by the dependent variable mean, where $\sigma_{\text{HDI}} = 0.13$ is the standard deviation of the health deficiency index in the hospital population.

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

Table A11: Effects on likelihood of being observed as single for men and women – Weighting

	Dependent variable: =1 if ever single at ages 18+		
	Baseline sample		Linked only
	(1) Women	(2) Men	(3) Men
	Panel A: Effects of hospital admission		
Patient	0.040***	0.040***	0.046***
	(0.011)	(0.011)	(0.014)
% effect	8.0	8.4	6.3
	Panel B: Effects of health deficiency index		
Health deficiency index	0.111***	0.098***	0.139***
	(0.038)	(0.038)	(0.047)
% effect (σ)	2.9	2.6	2.5
Mean of Y	0.500	0.480	0.730
N	7,474	7,596	3,664

Notes: This table displays sibling fixed effects estimates from regressions that re-weight the data by the following observable characteristics in the hospital records: admission age, admission year, and location of residence. Re-weighting the data ensures that each linked sample matches the proportions in the population of inpatient hospital admissions. The weighting procedure is described in more detail in Online Appendix D. In panel A, the explanatory variable of interest is an indicator for hospital admission, and in panel B, it is the health deficiency index. The estimates correspond to the outcomes variables and samples used in Table 3. See Table 3 for a description of the control variables. Standard errors are clustered by childhood household. Percent effects in panel B are computed as $\beta \times \sigma_{\mathrm{HDI}}$ scaled by the dependent variable mean, where $\sigma_{\mathrm{HDI}} = 0.13$ is the standard deviation of the health deficiency index in the hospital population.

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

Table A12: Long-run outcomes: Robustness to selective mortality

	(1) Baseline estimate	(2) Drop high mortality	(3) Drop infant admission	(4) Drop multiple admissions	(5) Drop low mortality	(6) Drop contagious
	Panel A: Effects on P(Class ↗)					
Health deficiency index	−0.098**	−0.135***	−0.091*	−0.116**	−0.097**	−0.141***
	(0.046)	(0.051)	(0.048)	(0.049)	(0.047)	(0.053)
Mean of Y	0.357	0.356	0.359	0.357	0.355	0.356
	Panel B: Effects on P(Class ↘)					
Health deficiency index	0.084*	0.083*	0.072	0.086*	0.091**	0.079
	(0.043)	(0.049)	(0.045)	(0.046)	(0.044)	(0.051)
Mean of Y	0.260	0.256	0.261	0.260	0.261	0.263
	Panel C: Effects on P(White collar)					
Health deficiency index	−0.142***	−0.164***	−0.158***	−0.171***	−0.146***	−0.181***
	(0.046)	(0.053)	(0.048)	(0.049)	(0.047)	(0.054)
Mean of Y	0.273	0.269	0.276	0.267	0.278	0.276
	Panel D: Effects on P(Skilled +)					
Health deficiency index	−0.116**	−0.149**	−0.099*	−0.126**	−0.118**	−0.134**
	(0.054)	(0.061)	(0.056)	(0.057)	(0.055)	(0.063)
Mean of Y	0.539	0.538	0.546	0.534	0.541	0.541
	Panel E: Effects on P(Unskilled)					
Health deficiency index	0.103***	0.110**	0.102**	0.108***	0.095**	0.126***
	(0.039)	(0.043)	(0.040)	(0.042)	(0.039)	(0.046)
Mean of Y	0.153	0.155	0.153	0.152	0.151	0.154
	Panel F: Effects on log occupational wage					
Health deficiency index	−0.104***	−0.121***	−0.120***	−0.131***	−0.102**	−0.142***
	(0.039)	(0.045)	(0.041)	(0.041)	(0.040)	(0.046)
Mean of Y	4.629	4.628	4.631	4.625	4.631	4.635
N	3,698	3,328	3,350	3,256	3,328	3,044

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

Table A13: Long-run outcomes: Robustness to sample selection

	(1) Baseline estimate	(2) Add multiple siblings	(3) Add multiple patient hhlds.	(4) County of London only	(5) Drop Guy's Hospital	(6) Unique within census county	(7) Hospital-census county match
Panel A: Effects on P(Class $\nearrow$)							
Health deficiency index	−0.098**	−0.069	−0.066	−0.107**	−0.095**	−0.111**	−0.119**
	(0.046)	(0.042)	(0.042)	(0.050)	(0.048)	(0.048)	(0.055)
Mean of Y	0.357	0.355	0.355	0.367	0.357	0.353	0.363
Panel B: Effects on P(Class $\searrow$)							
Health deficiency index	0.084*	0.078*	0.072*	0.084*	0.097**	0.064	0.073
	(0.043)	(0.040)	(0.040)	(0.046)	(0.046)	(0.045)	(0.053)
Mean of Y	0.260	0.259	0.260	0.265	0.261	0.258	0.267
Panel C: Effects on P(White collar)							
Health deficiency index	−0.142***	−0.120***	−0.117***	−0.141***	−0.152***	−0.153***	−0.167***
	(0.046)	(0.043)	(0.042)	(0.050)	(0.049)	(0.049)	(0.057)
Mean of Y	0.273	0.275	0.274	0.278	0.279	0.276	0.286
N	3,698	4,473	4,512	3,032	3,332	3,302	2,366
Panel D: Effects on P(Skilled +)							
Health deficiency index	−0.116**	−0.089*	−0.081	−0.125**	−0.145**	−0.122**	−0.132**
	(0.054)	(0.050)	(0.050)	(0.059)	(0.058)	(0.057)	(0.066)
Mean of Y	0.539	0.539	0.539	0.546	0.550	0.540	0.550
Panel E: Effects on P(Unskilled)							
Health deficiency index	0.103***	0.070*	0.071**	0.119***	0.099**	0.084**	0.073
	(0.039)	(0.036)	(0.036)	(0.042)	(0.041)	(0.041)	(0.048)
Mean of Y	0.153	0.156	0.155	0.144	0.145	0.154	0.142
Panel F: Effects on log occupational wage							
Health deficiency index	−0.104***	−0.083**	−0.083**	−0.095**	−0.114***	−0.113***	−0.115**
	(0.039)	(0.037)	(0.036)	(0.042)	(0.042)	(0.042)	(0.048)
Mean of Y	4.629	4.631	4.631	4.641	4.635	4.630	4.651
N	3,698	4,473	4,512	3,032	3,332	3,302	2,366

Notes: This table is identical to Table 6 with the exception that the treatment variable is changed from an indicator for hospitalization to the continuous health deficiency index. Standard errors are clustered by childhood household.

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

Table A14: Long-run outcomes: Robustness to variations in occupational status

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

Notes: This table is identical to Table 7 with the exception that the treatment variable is changed from an indicator for hospitalization to the continuous health deficiency index. Standard errors are clustered by childhood household.

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

Table A15: Robustness of effects on likelihood of being enumerated as a single adult

	(1) Drop high mortality	(2) Drop infant admission	(3) Drop multiple admissions	(4) Drop low mortality	(5) Drop contagious	(6) Add multiple siblings	(7) Add multiple patient hhlds.	(8) County of London only	(9) Drop Guy's Hospital	(10) Unique within census county	(11) Hospital-census county match
	Panel A: Effects on P(Ever single marital status) for women in baseline sample										
Patient	0.042***	0.040***	0.040***	0.038***	0.040***	0.042***	0.039***	0.051***	0.038***	0.042***	0.056***
	(0.012)	(0.012)	(0.012)	(0.012)	(0.013)	(0.011)	(0.011)	(0.013)	(0.012)	(0.012)	(0.014)
Health deficiency index	0.133***	0.104***	0.114***	0.101***	0.106**	0.114***	0.105***	0.130***	0.112***	0.117***	0.155***
	(0.044)	(0.039)	(0.039)	(0.038)	(0.046)	(0.036)	(0.038)	(0.042)	(0.040)	(0.039)	(0.046)
Mean of Y	0.497	0.497	0.497	0.493	0.505	0.490	0.498	0.489	0.502	0.500	0.501
N	6,756	6,846	6,770	6,724	5,952	9,201	7,579	5,890	6,790	6,806	4,948
	Panel B: Effects on P(Ever single marital status) for men in baseline sample										
Patient	0.043***	0.038***	0.041***	0.038***	0.040***	0.037***	0.039***	0.052***	0.034***	0.032***	0.043***
	(0.012)	(0.012)	(0.012)	(0.012)	(0.013)	(0.011)	(0.011)	(0.013)	(0.012)	(0.012)	(0.014)
Health deficiency index	0.113***	0.087**	0.092**	0.106***	0.086*	0.088**	0.089**	0.134***	0.074*	0.064	0.099**
	(0.043)	(0.040)	(0.040)	(0.038)	(0.045)	(0.035)	(0.037)	(0.041)	(0.040)	(0.040)	(0.047)
Mean of Y	0.479	0.480	0.481	0.477	0.479	0.472	0.478	0.477	0.481	0.480	0.478
N	6,806	6,806	6,830	6,844	6,144	9,317	7,746	6,098	6,864	6,816	5,050
	Panel C: Effects on P(Ever single marital status) for men in linked sample										
Patient	0.051***	0.047***	0.051***	0.048***	0.049***	0.030**	0.046***	0.059***	0.040***	0.036**	0.052***
	(0.015)	(0.015)	(0.015)	(0.015)	(0.016)	(0.013)	(0.014)	(0.016)	(0.015)	(0.015)	(0.018)
Health deficiency index	0.177***	0.138***	0.151***	0.137***	0.161***	0.083*	0.136***	0.173***	0.123**	0.099**	0.149**
	(0.055)	(0.050)	(0.050)	(0.048)	(0.058)	(0.043)	(0.047)	(0.052)	(0.050)	(0.050)	(0.059)
Mean of Y	0.723	0.722	0.727	0.722	0.722	0.723	0.721	0.716	0.726	0.725	0.711
N	3,304	3,282	3,280	3,286	2,992	4,631	3,922	2,944	3,320	3,280	2,462

Notes: Each cell comes from a separate sibling fixed effects regression. In Panels A to C, the dependent variables and estimation samples are identical to those presented in columns 1 to 3, respectively, of Table 3. Each column represents a separate modification to the main estimation sample. The exercises in columns 1 to 5 are identical to those presented in Table 5, while the exercises in columns 6 to 11 correspond to Table 6.

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

Table A16: Robustness of participation in schooling outcome

	(1) Column (3) Table 4	(2) Drop high mortality	(3) Drop infant admissions	(4) Drop multiple admissions	(5) Drop low mortality	(6) Drop contagious
	Panel A: Selective mortality					
Patient	−0.026***	−0.023***	−0.025***	−0.029***	−0.025***	−0.018**
	(0.008)	(0.009)	(0.009)	(0.009)	(0.009)	(0.009)
Health deficiency index	−0.079***	−0.070**	−0.079***	−0.091***	−0.081***	−0.047
	(0.028)	(0.032)	(0.030)	(0.031)	(0.029)	(0.031)
Mean of Y	0.743	0.755	0.750	0.751	0.745	0.748
N	3,040	2,728	2,684	2,672	2,698	2,504
	(7) Add multiple siblings	(8) Add multiple patient hhlds.	(9) County of London only	(10) Drop Guy's Hospital	(11) Unique within census county	(12) Hospital-census county match
	Panel B: Sample selection					
Patient	−0.028***	−0.027***	−0.024**	−0.030***	−0.019**	−0.017*
	(0.008)	(0.008)	(0.010)	(0.009)	(0.008)	(0.010)
Health deficiency index	−0.083***	−0.080***	−0.070**	−0.092***	−0.068**	−0.061*
	(0.026)	(0.026)	(0.030)	(0.031)	(0.029)	(0.033)
Mean of Y	0.744	0.745	0.732	0.744	0.748	0.726
N	3,340	3,363	2,418	2,774	2,784	1,980

Notes: Each cell comes from a separate sibling fixed effects regression. Column 1 of panel A reproduces the estimate for the schooling outcome from column 3 of Table 4, which is based on a sample of individuals aged 5 to 10 when enumerated in the census. See Table 4 for a list of variables included in the regressions. See Tables 5 and 6 for a description of the sample restrictions in the remaining columns and panels. Standard errors are clustered by childhood household.

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

Table A17: Robustness of pre-existing disability outcome

	(1) Column (6) Table 4	(2) Drop high mortality	(3) Drop infant admissions	(4) Drop multiple admissions	(5) Drop low mortality	(6) Drop contagious
			Panel A: Selective mortality			
Patient	0.101	0.104	0.049	0.067	0.102	0.198
	(0.118)	(0.129)	(0.113)	(0.099)	(0.132)	(0.140)
Health deficiency index	0.029	−0.023	−0.082	−0.060	0.123	0.235
	(0.303)	(0.432)	(0.304)	(0.269)	(0.320)	(0.435)
Mean of Y	0.154	0.171	0.151	0.140	0.160	0.171
N	9,096	8,194	8,612	7,874	8,128	7,018
	(7) Add multiple siblings	(8) Add multiple patient hhlds.	(9) County of London only	(10) Drop Guy's Hospital	(11) Unique within census county	(12) Hospital-census county match
			Panel B: Sample selection			
Patient	0.101	0.100	0.072	0.134	0.048	0.044
	(0.116)	(0.113)	(0.127)	(0.132)	(0.124)	(0.158)
Health deficiency index	0.054	0.059	−0.082	0.095	−0.130	−0.239
	(0.280)	(0.271)	(0.328)	(0.360)	(0.324)	(0.422)
Mean of Y	0.153	0.150	0.151	0.161	0.147	0.204
N	12,420	12,637	7,268	8,058	8,174	5,892

Notes: Each cell comes from a separate sibling fixed effects regression. Column 1 of panel A reproduces the estimate for the pre-existing disability outcome from column 6 of Table 4. See Table 4 for a list of variables included in the regressions. See Tables 5 and 6 for a description of the sample restrictions in the remaining columns and panels. Standard errors are clustered by childhood household.

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

Table A18: Robustness of childhood disability outcome

	(1) Column (9) Table 4	(2) Drop high mortality	(3) Drop infant admissions	(4) Drop multiple admissions	(5) Drop low mortality	(6) Drop contagious
	Panel A: Selective mortality					
Patient	0.746***	0.750***	0.828***	0.705***	0.719***	0.726***
	(0.185)	(0.200)	(0.208)	(0.193)	(0.197)	(0.210)
Health deficiency index	2.235***	2.392***	2.325***	1.955***	2.148***	2.141***
	(0.569)	(0.670)	(0.615)	(0.559)	(0.599)	(0.670)
Mean of Y	0.513	0.531	0.545	0.474	0.542	0.535
N	11,312	10,178	10,098	9,912	10,154	9,160
	(7) Add multiple siblings	(8) Add multiple patient hhlds.	(9) County of London only	(10) Drop Guy's Hospital	(11) Unique within census county	(12) Hospital-census county match
	Panel B: Sample selection					
Patient	0.716***	0.713***	0.573***	0.794***	0.761***	0.974***
	(0.165)	(0.161)	(0.196)	(0.195)	(0.194)	(0.254)
Health deficiency index	2.170***	2.147***	1.772***	2.277***	2.287***	2.899***
	(0.543)	(0.528)	(0.607)	(0.597)	(0.600)	(0.773)
Mean of Y	0.422	0.420	0.454	0.529	0.526	0.671
N	16,346	16,656	9,030	10,202	10,270	7,454

Notes: Each cell comes from a separate sibling fixed effects regression. Column 1 of panel A reproduces the estimate for the childhood disability outcome from column 9 of Table 4. See Table 4 for a list of variables included in the regressions. See Tables 5 and 6 for a description of the sample restrictions in the remaining columns and panels. Standard errors are clustered by childhood household.

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

Table A19: Robustness of long-run disability outcomes

	(1) Column (10) Table 4	(2) Drop high mortality	(3) Drop infant admissions	(4) Drop multiple admissions	(5) Drop low mortality	(6) Drop contagious
	Panel A: Selective mortality					
Patient	0.739***	0.853***	0.730***	0.701***	0.764***	0.920***
	(0.233)	(0.248)	(0.239)	(0.253)	(0.252)	(0.250)
Health deficiency index	1.866**	2.815***	1.674**	1.783**	1.949**	3.106***
	(0.799)	(0.897)	(0.791)	(0.897)	(0.805)	(0.908)
Mean of Y	0.580	0.593	0.562	0.554	0.593	0.567
N	4,312	3,880	3,918	3,790	3,880	3,530
	(7) Add multiple siblings	(8) Add multiple patient hhlds.	(9) County of London only	(10) Drop Guy's Hospital	(11) Unique within census county	(12) Hospital-census county match
	Panel B: Sample selection					
Patient	0.560**	0.559**	0.651**	0.841***	0.815***	1.042***
	(0.240)	(0.237)	(0.260)	(0.254)	(0.238)	(0.308)
Health deficiency index	1.416*	1.408*	1.412	2.049**	2.164***	2.741***
	(0.818)	(0.806)	(0.863)	(0.883)	(0.714)	(0.949)
Mean of Y	0.484	0.479	0.569	0.593	0.544	0.576
N	4,547	4,591	3,512	3,878	3,862	2,776

Notes: Each cell comes from a separate sibling fixed effects regression. Column 1 of panel A reproduces the estimate for the long-run disability outcome from column 10 of Table 4. See Table 4 for a list of variables included in the regressions. See Tables 5 and 6 for a description of the sample restrictions in the remaining columns and panels. Standard errors are clustered by childhood household.

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

Table A20: Sibling-specific determinants of hospitalization

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

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

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

Table A21: Patient health deficiency index and likelihood of linkage to census

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

Notes: This table presents OLS estimates from specifications in which the dependent variable is an indicator for a unique match from the hospital records to a census. The only explanatory variable is the health deficiency index, the coefficients for which are reported in the table. Columns 1 to 3 present results for linkages from the hospital records to the 1881, 1891, and 1901 censuses, respectively. Column 4 pools together the samples in the preceding columns. The samples consists of all patients admitted to the hospital at ages 0 to 11 years old no more than 10 years prior to the census enumeration date and discharged from the hospital prior to the census enumeration date.
Point estimates marked ***, **, and * are statistically significant at the 1, 5, and 10 percent levels, respectively.

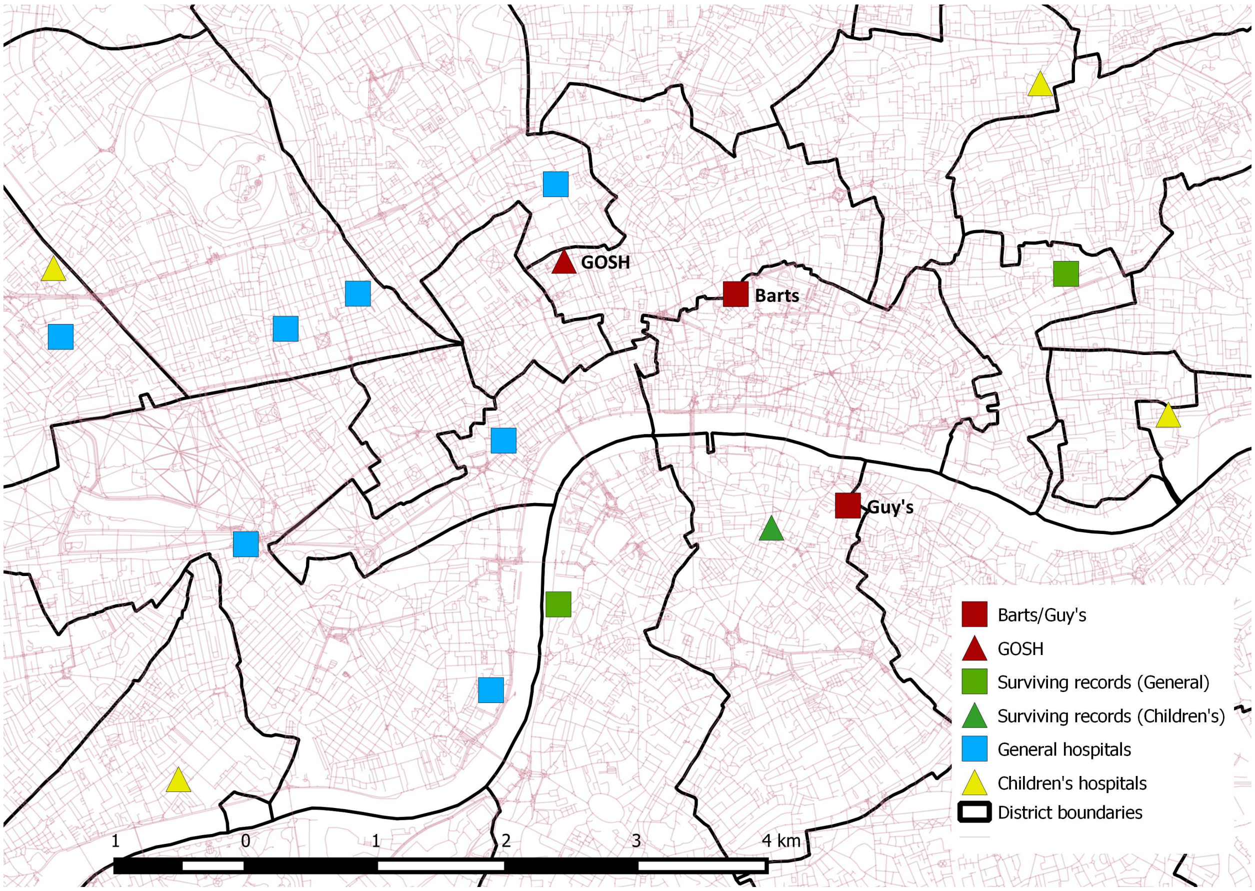
Table A22: Census-to-census linkage rates

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

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

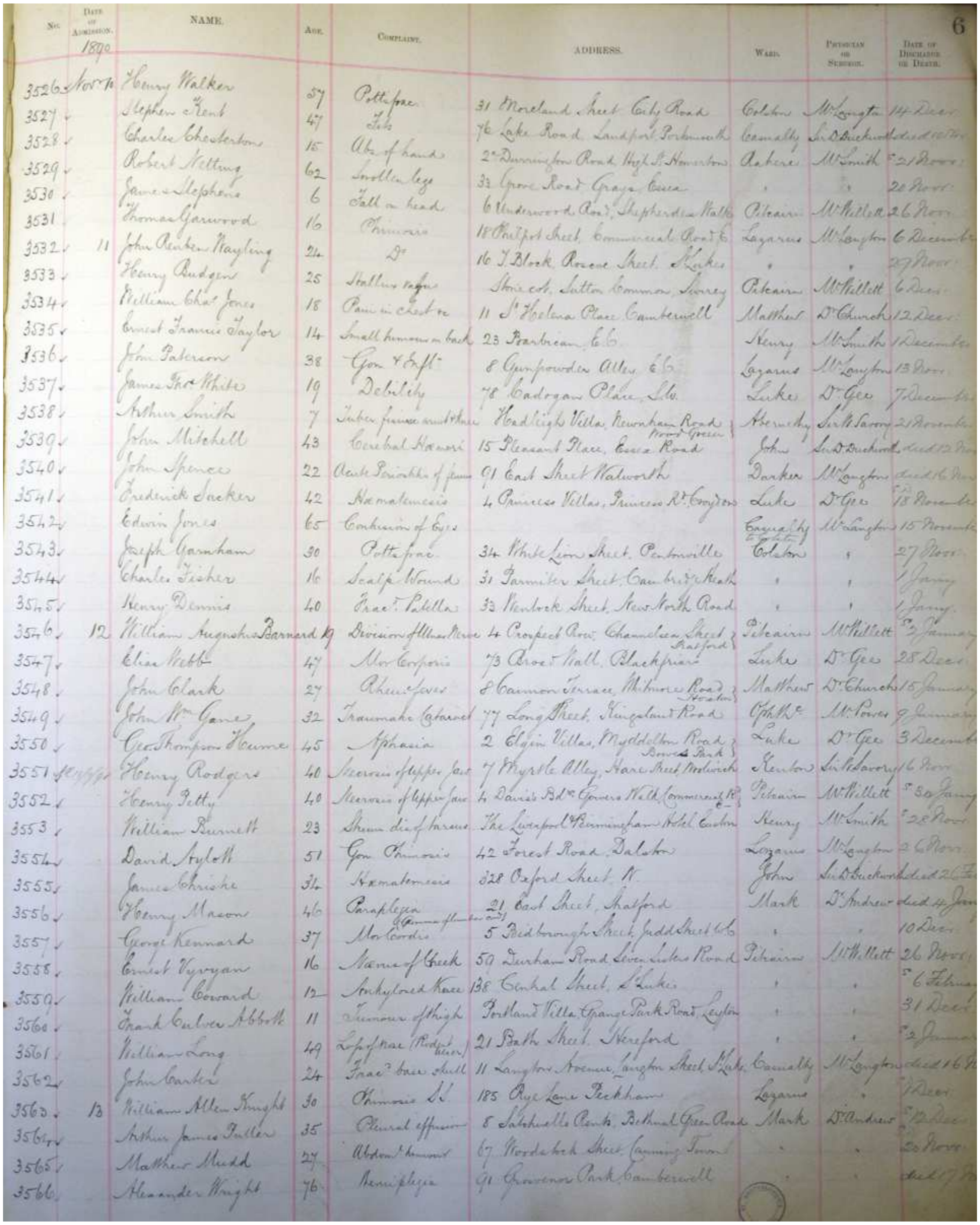
F Appendix Figures

Figure A1: London hospital locations and surviving records



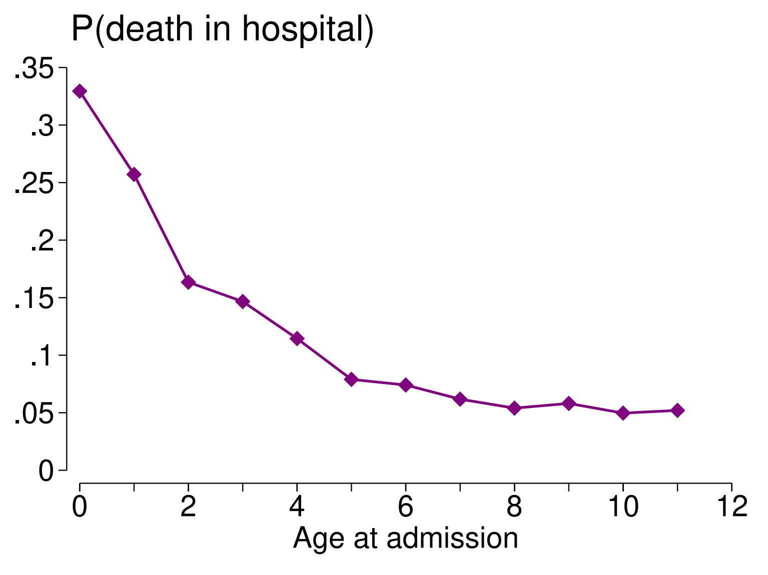
Notes: A map of central London marking the locations of hospitals in the empirical sample (red squares and triangle), general hospitals (squares symbol), and children's hospitals (triangle symbol). The subset of these hospitals with surviving archival records is marked in green.

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

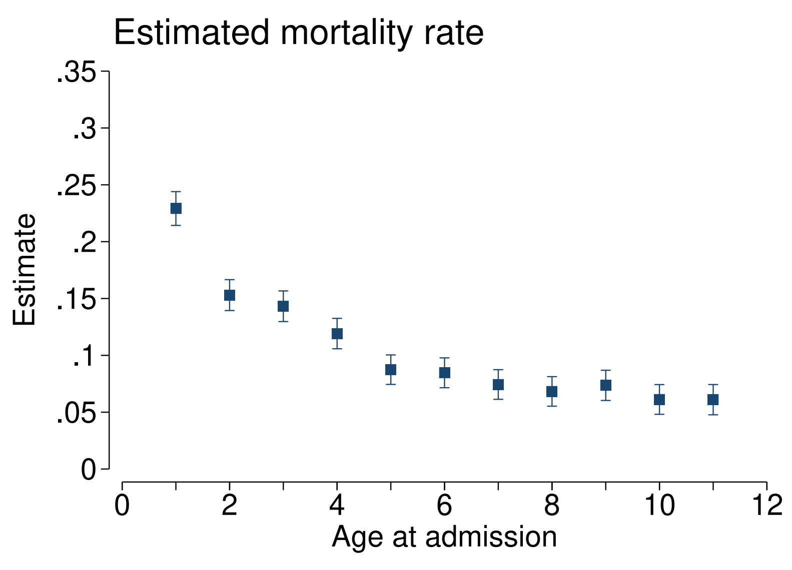


Notes: Sample page of an inpatient admission register from November 1890 for St. Bartholomew's Hospital in London. Each page contains the date of admission, the patient's name, age, complaint and address, the name of the ward in which the patient was admitted, the name of the physician or surgeon who treated the patient, and the date of discharge or death. Source: Photograph taken by authors courtesy of Barts Health NHS Trust Archives. Document reference number: SBHB/MR/2/6.

Figure A3: In-hospital mortality by age at admission



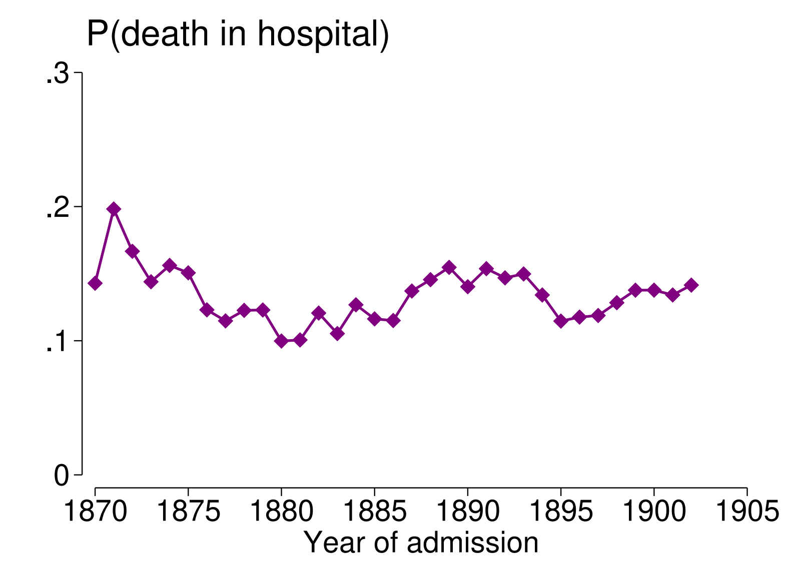
(a) Raw data



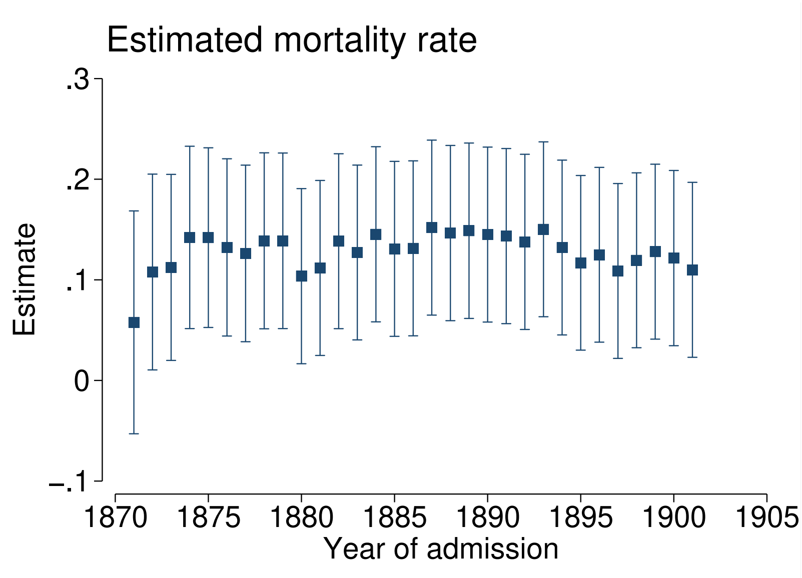
(b) Regression adjusted

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

Figure A4: In-hospital mortality by year of admission



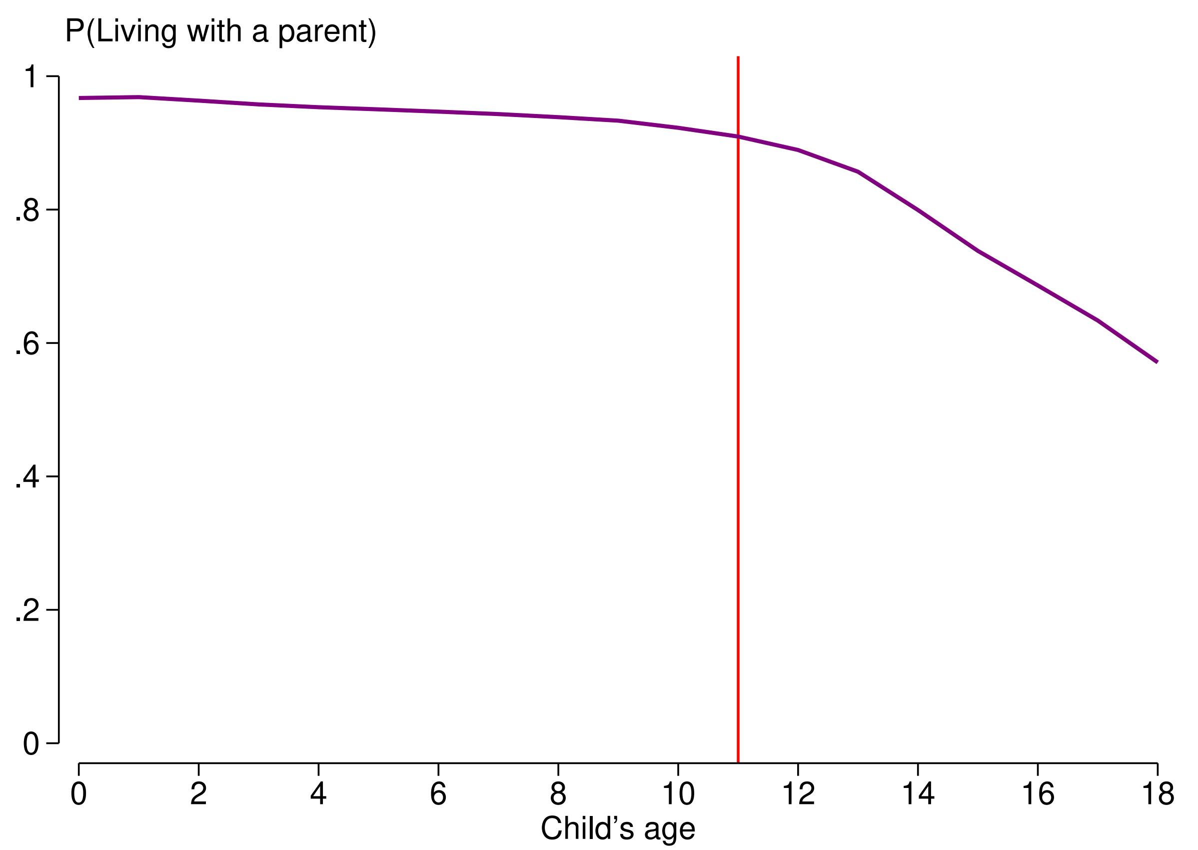
(a) Raw data



(b) Regression adjusted

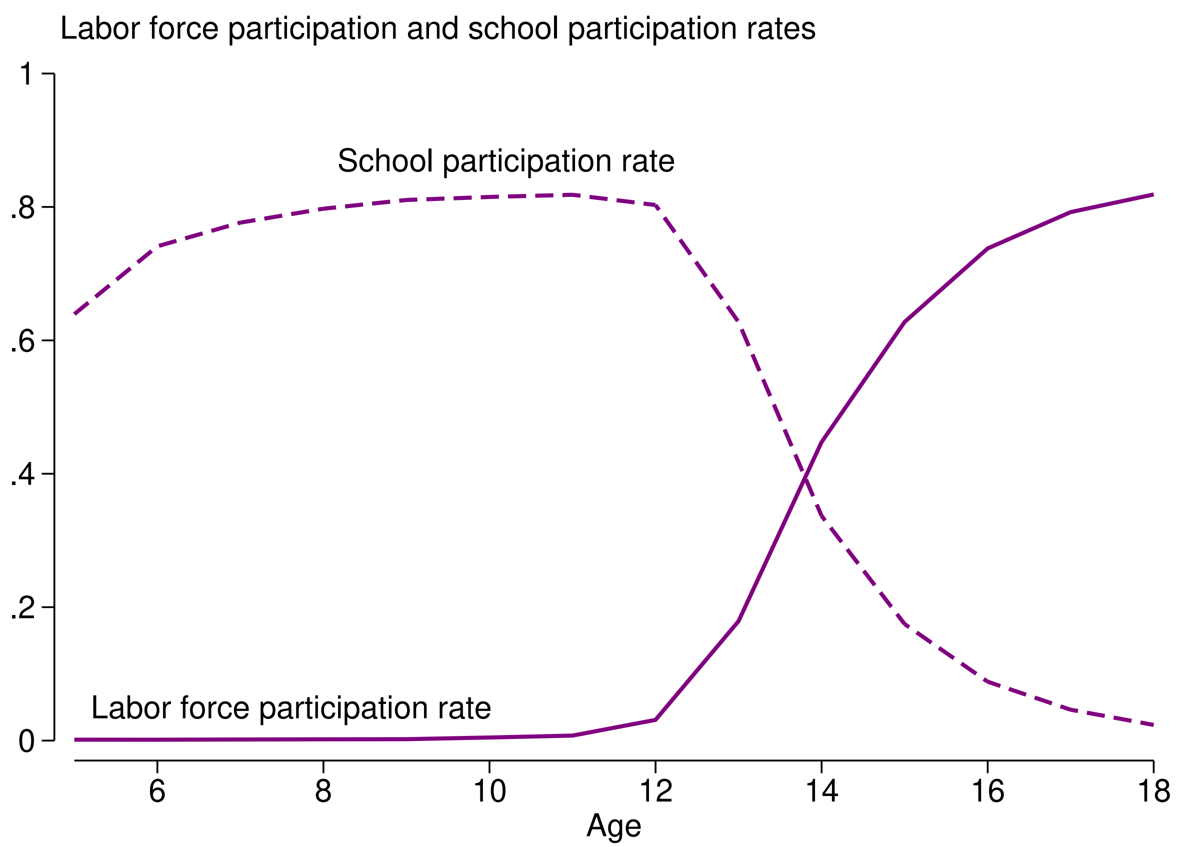
Notes: Panel A plots the in-hospital mortality rate by year of admission and Panel B presents regression-adjusted estimates. Panel B plots estimated fixed effects on year of admission (with 1870 as the excluded category) from a linear probability model which also includes admission age, hospital, gender,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

Figure A5: Share of children living with a parent in 1881, by age



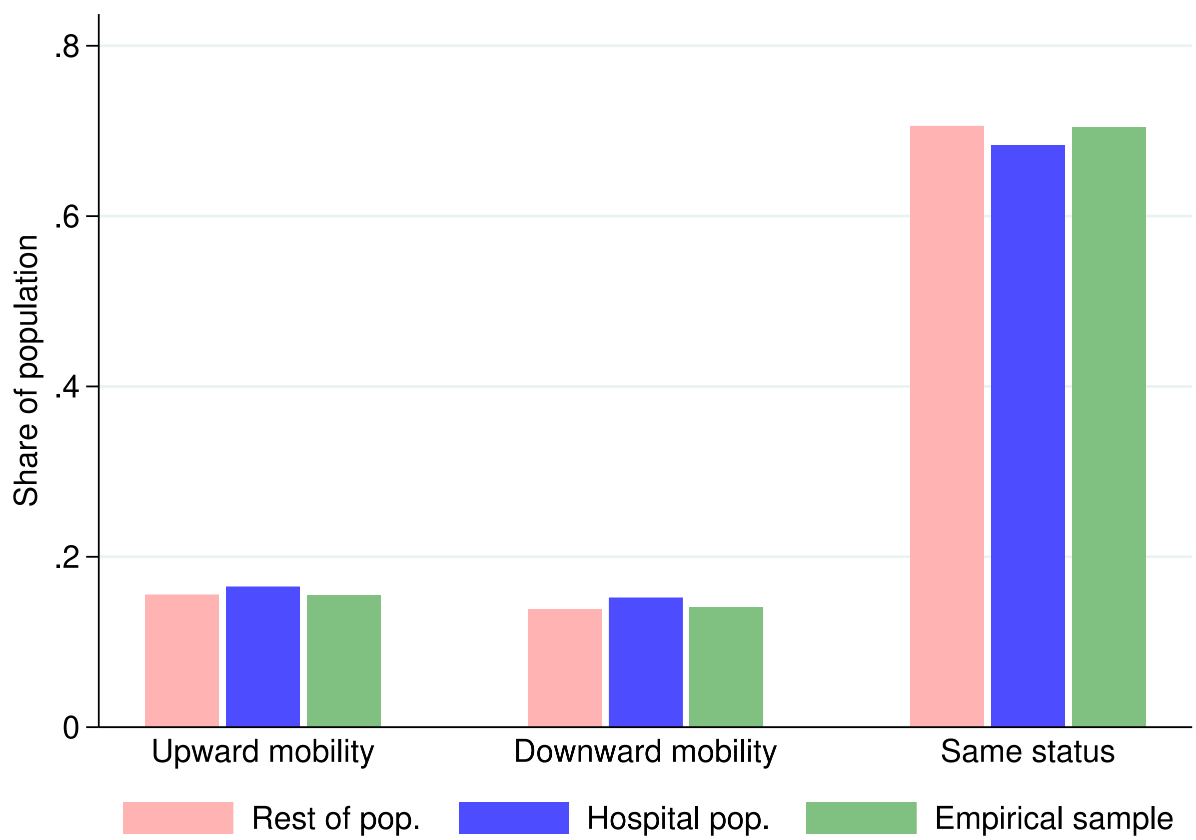
Notes: This figure plots the share of children who were enumerated in the 1881 census in a household in which at least one parent was present. The sample consists of all households in the County of London. Results are similar in the 1891 census. The vertical red line indicates age 11, which is the oldest age at admission of a hospital patient in our data.

Figure A6: Labor force participation and participation in schooling by age in 1881



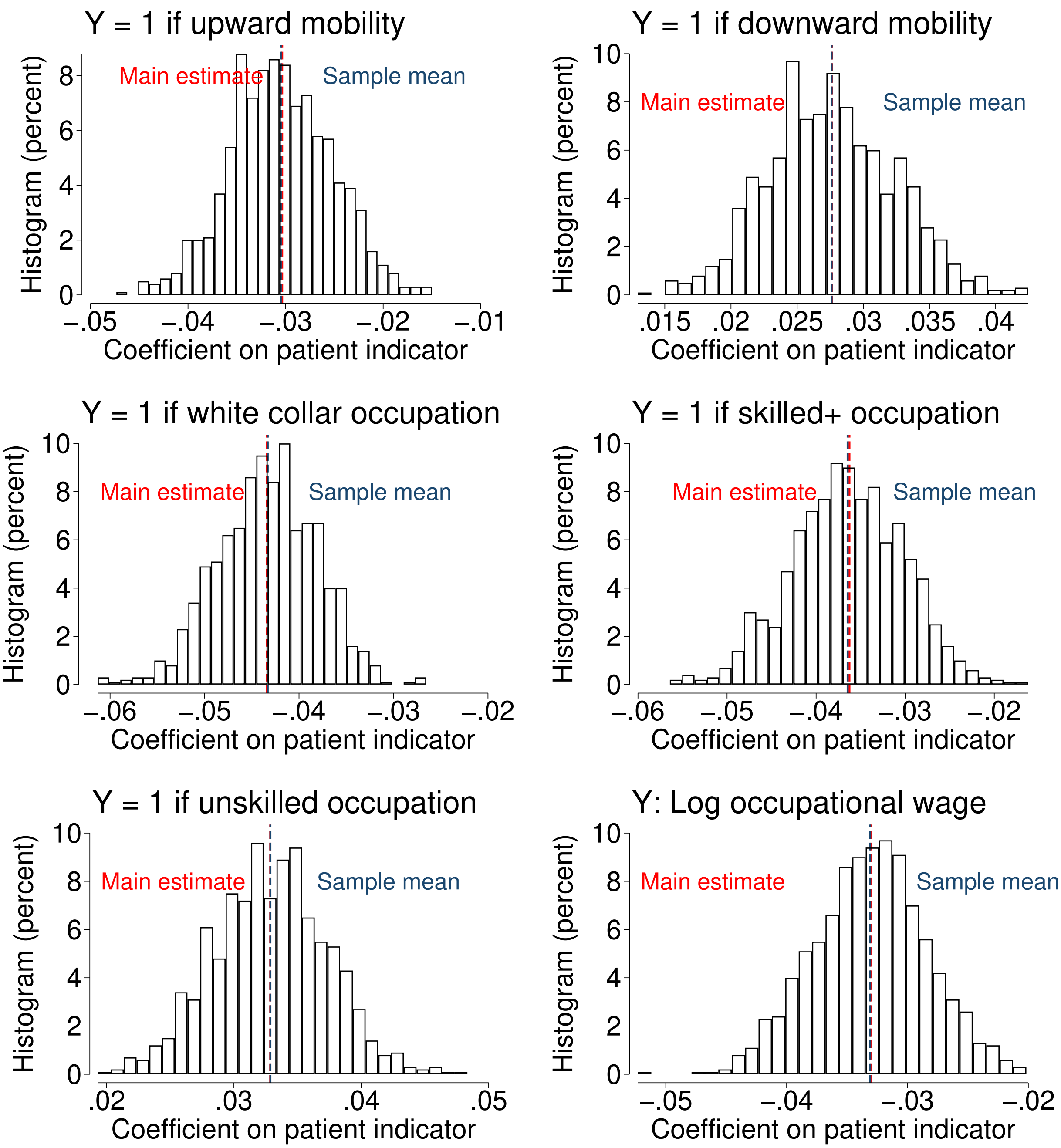
Notes: This figure plots the labor force participation rate (solid line) and the rate of participation in schooling (dashed line) by age (5 to 18) in the 1881 Population Census of England, for male individuals residing in the County of London. An individual is considered to be in the labor force if any gainful occupation is reported in the census as measured by a valid Historical International Standard Classification of Occupations (HISCO) code. See Section III.F for a description of how participation in schooling is coded. Results are similar in the 1891 census.

Figure A7: Changes in father's occupation across adjacent censuses



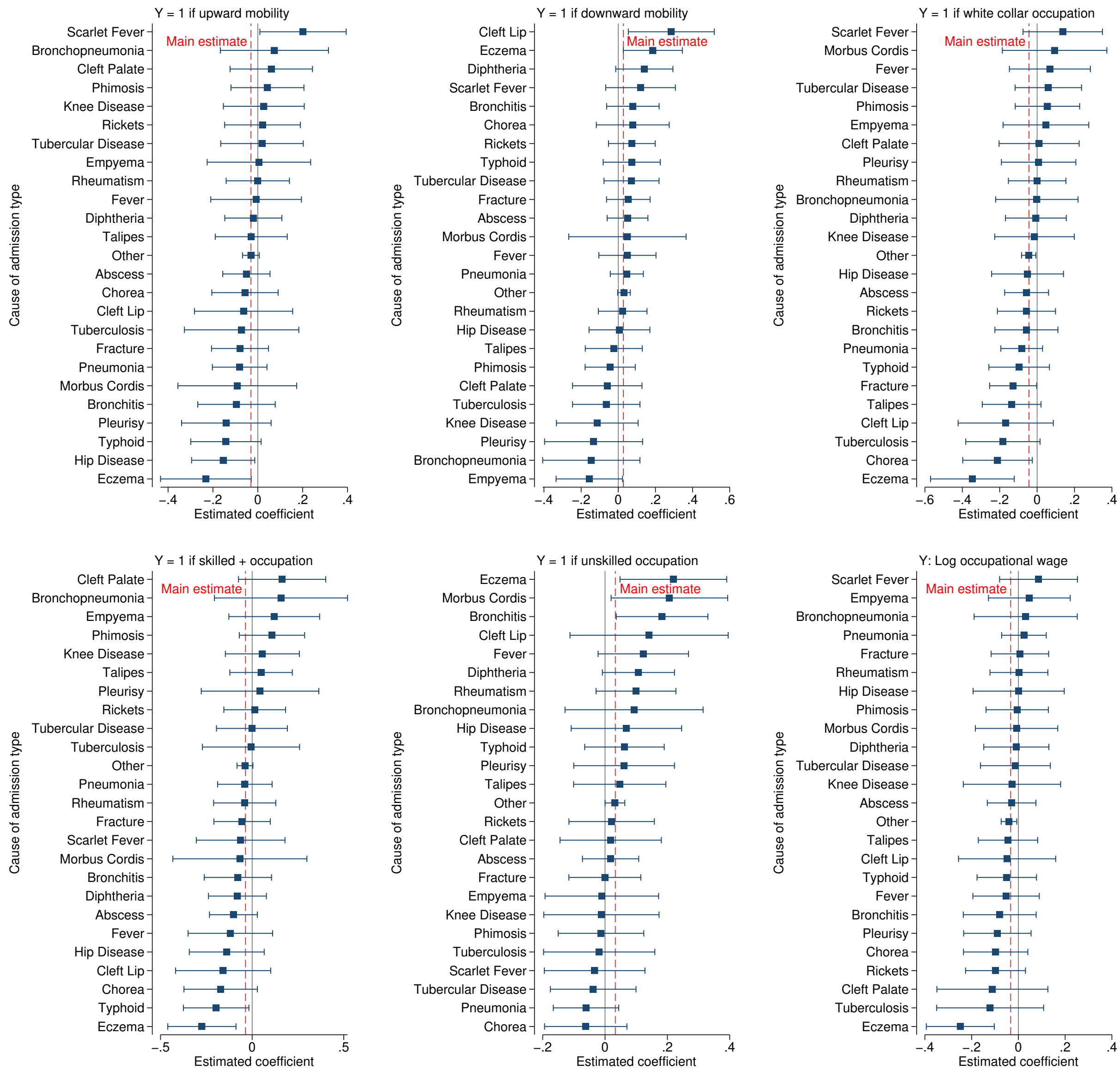
Notes: This figure presents summary measures of occupational changes for fathers linked between the 1881 and 1891 or 1891 and 1901 censuses. It shows the share of fathers whose occupational rank increases (upward mobility), decreases (downward mobility) or stays the same (same status) from one census to the next. Shares are shown separately for the subset of the main empirical sample with linked fathers (green), the rest of the hospital population (blue), and the rest of the overall population (pink). See Online Appendix A.8 for a description of the procedure for linking fathers.

Figure A8: Randomly assign lower father's SES to 15% of patients 1000 times



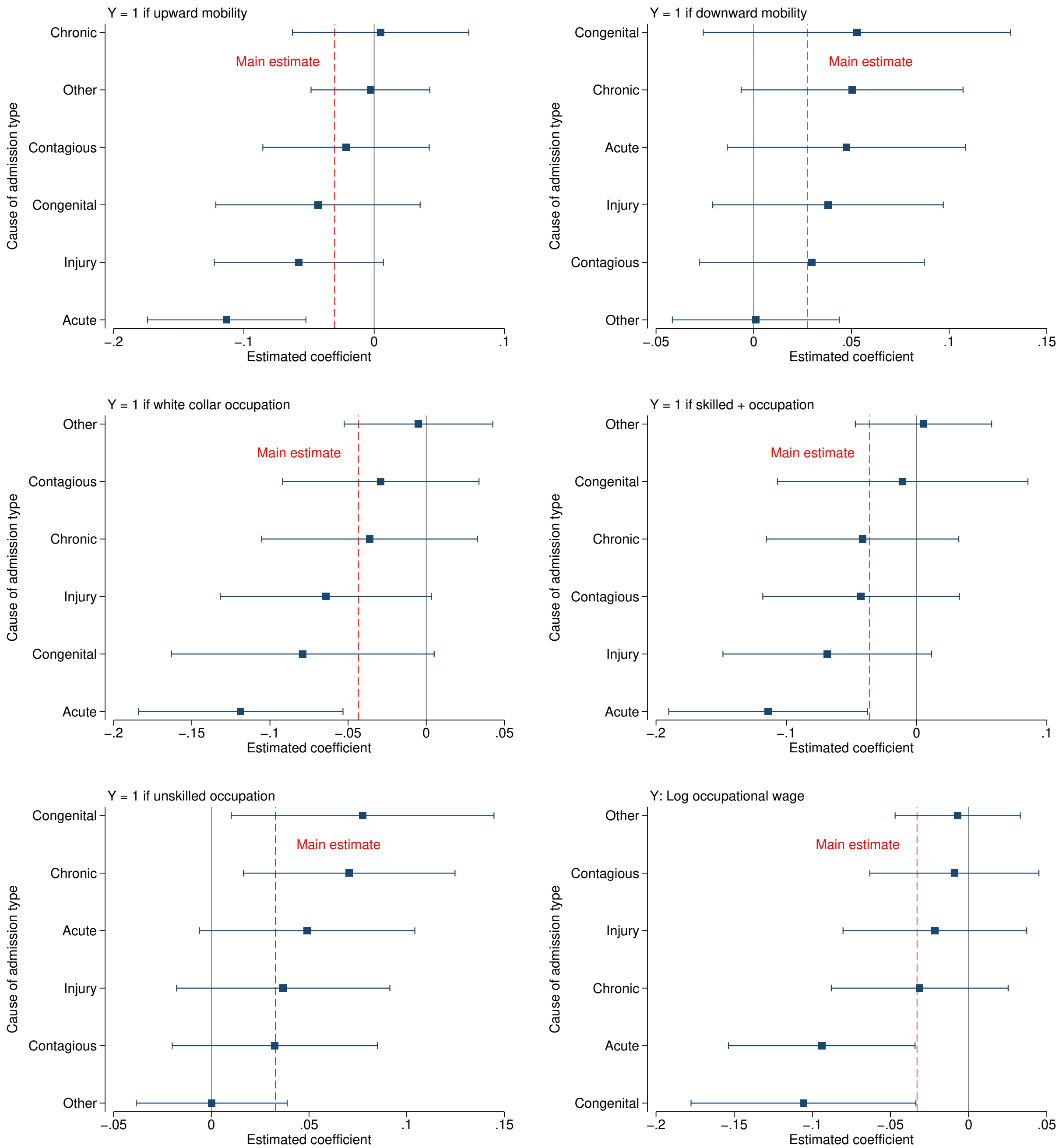
Notes: Each figure plots the distribution of estimated coefficients on the patient indicator from 1000 iterations of assigning the father's socioeconomic status (SES) to be one rank lower for 15 percent of the patients in the sample and re-estimating the main specification for the six dependent variables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

Figure A9: Effects of hospitalization by common causes of admission



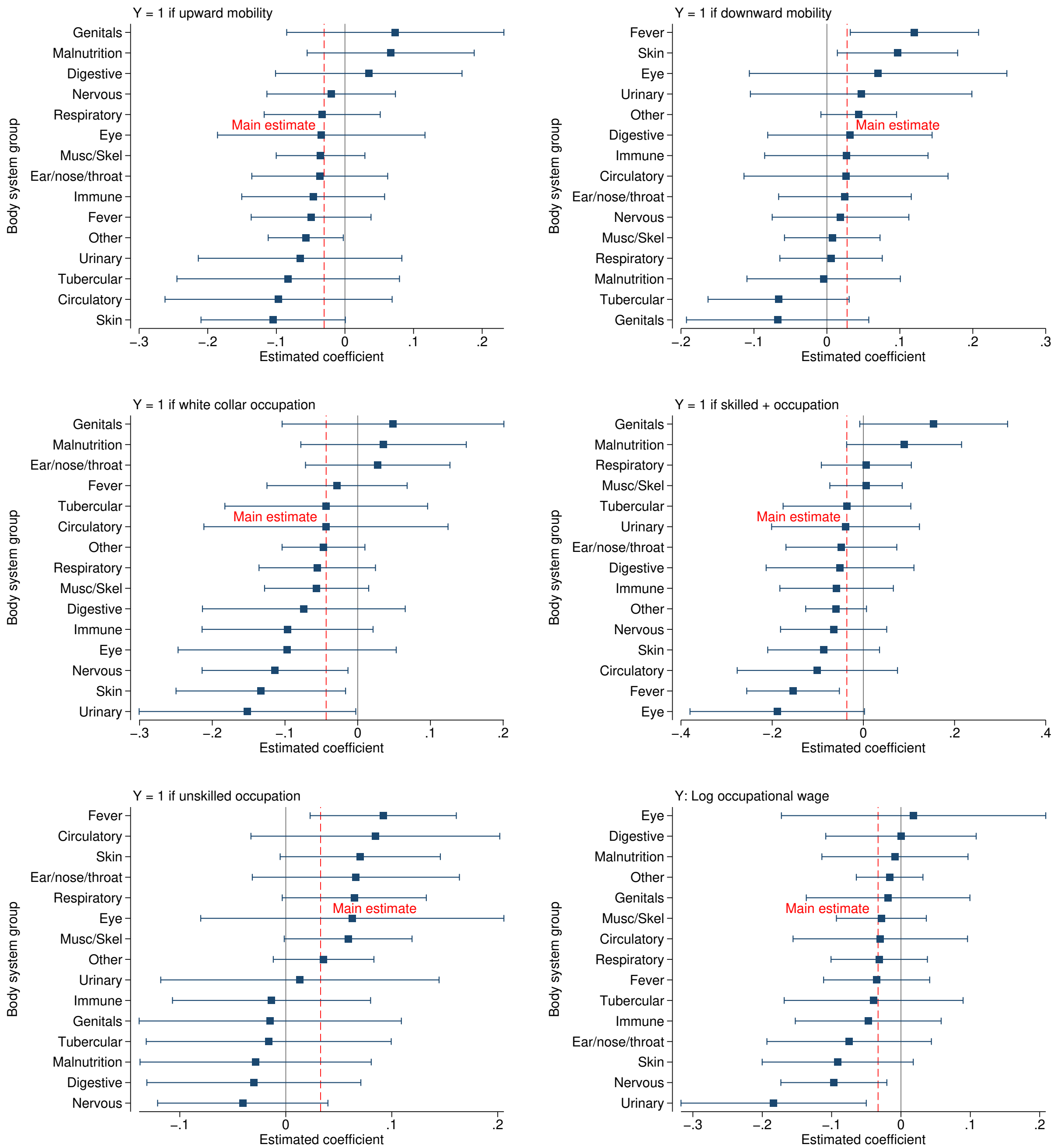
Notes: Each figure plots estimated coefficients on indicators for the most common causes of admission listed in Table A2, and 95-percent confidence intervals from a single sibling fixed effects regression with one of six long-run occupational outcome variables shown in Table 2. In addition to the causes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

Figure A10: Effects of hospitalization by hospital admission type



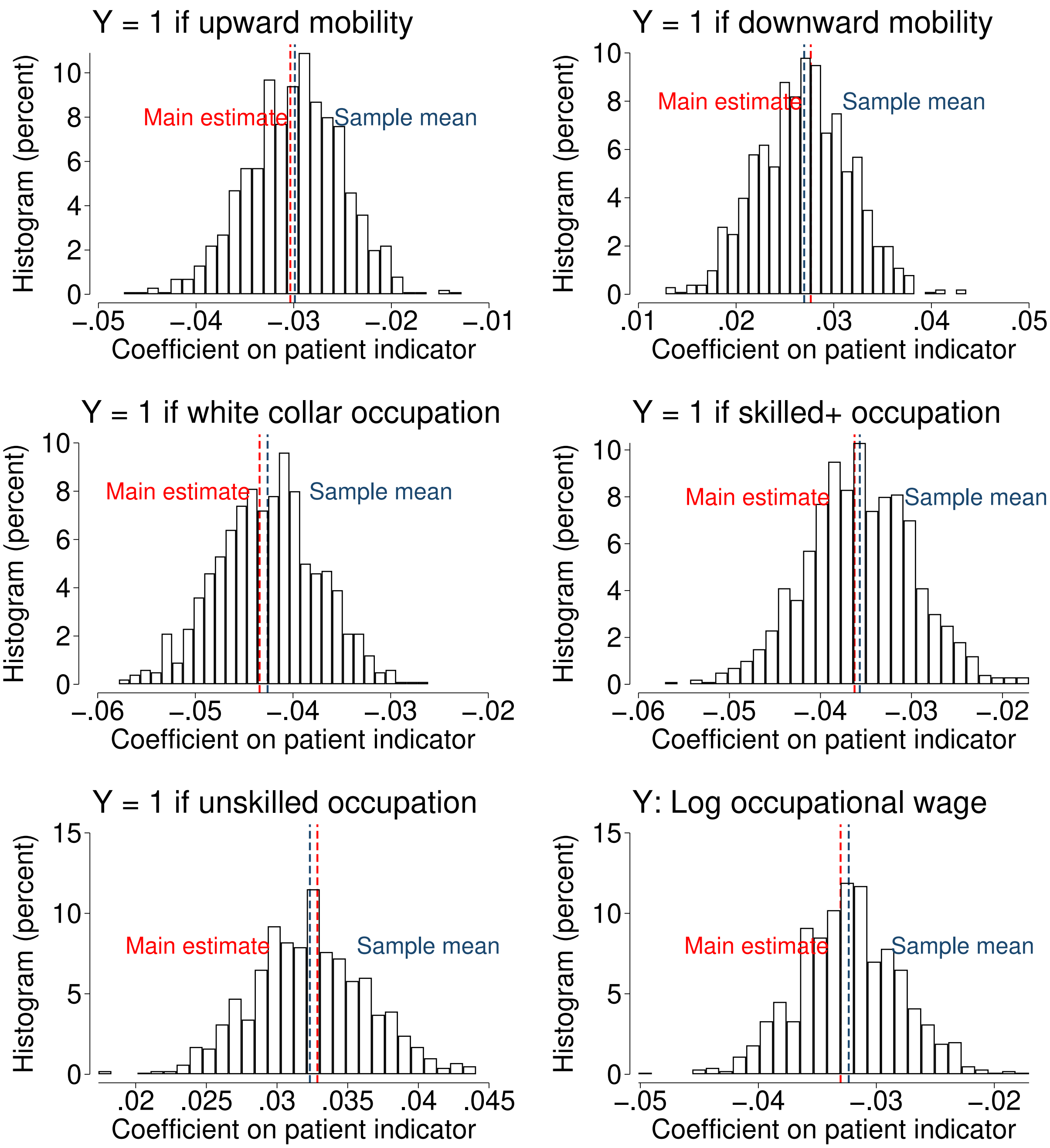
Notes: Each figure plots the estimated coefficients on indicators for a type of hospital admission and 95-percent confidence intervals from a single sibling fixed effects regression with one of six long-run occupational outcome variables shown in Table 2. In addition to the causes of admission plotted,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

Figure A11: Effects of hospitalization by body system classification



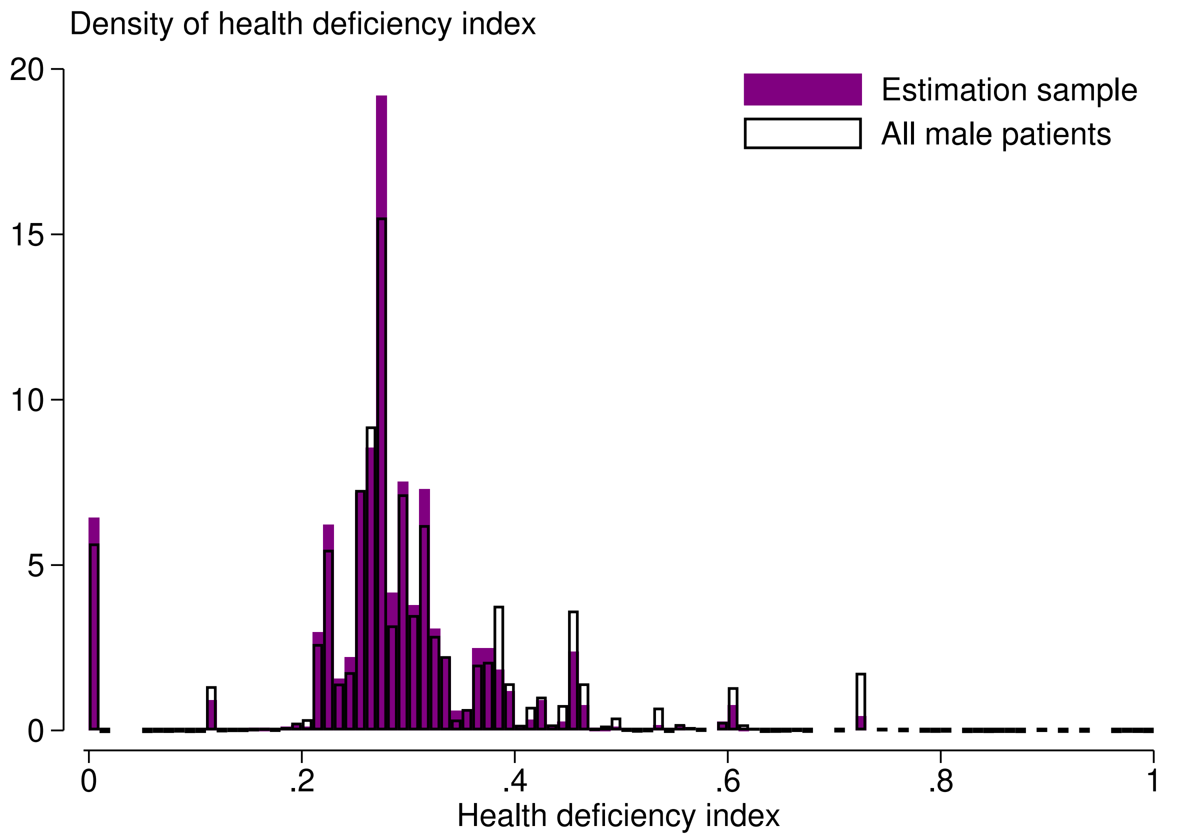
Notes: Each figure plots the estimated coefficients on indicators for the type of hospital admission, classified by body system, and 95-percent confidence intervals from a single sibling fixed effects regression with one of six long-run occupational outcome variables shown in Table 2. In addition to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

Figure A12: Randomly assign treatment (=1) to 10% of siblings 1000 times



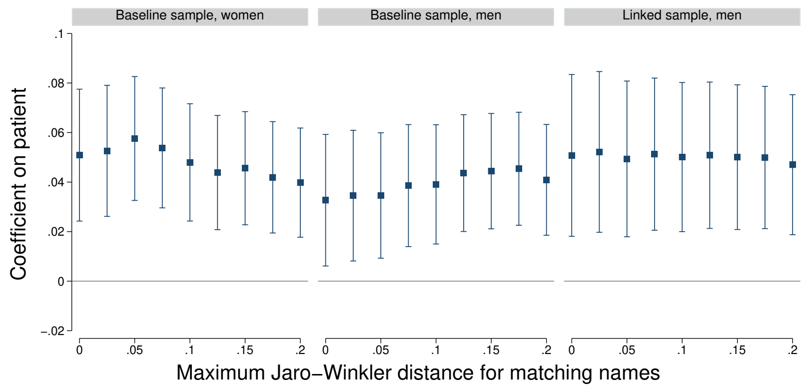
Notes: Each figure plots the distribution of estimated coefficients on the hospitalization indicator from 1,000 iterations of assigning the treatment indicator equal to one for 10 percent of the non-hospitalized siblings in the sample and re-estimating the main specification for the six dependent variables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

Figure A13: Density of health deficiency index in population vs. estimation sample

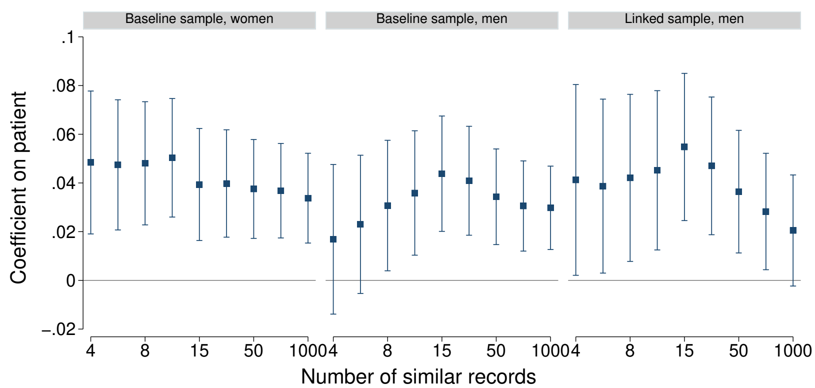


Notes: This figure presents a histogram of the health deficiency index for the population of male patients admitted to the hospitals in the full sample (white) and for the patients in the final estimation sample (solid purple). See Online Appendix C for a description of the procedure used to construct the health deficiency index.

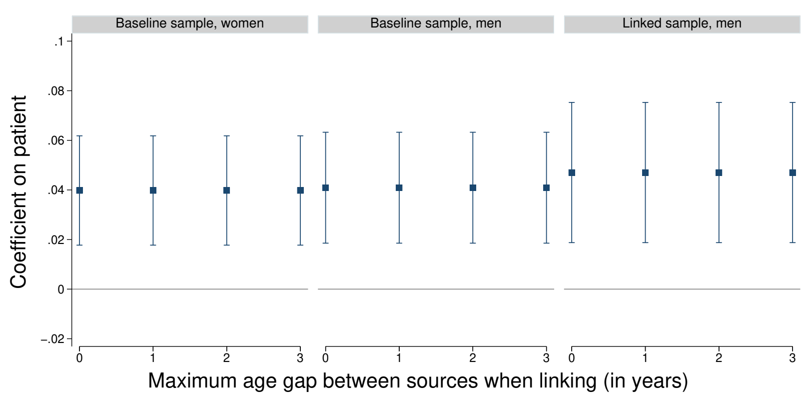
Figure A14: Robustness of effects on likelihood of being enumerated as a single adult



(a) Changing Jaro-Winkler distance threshold



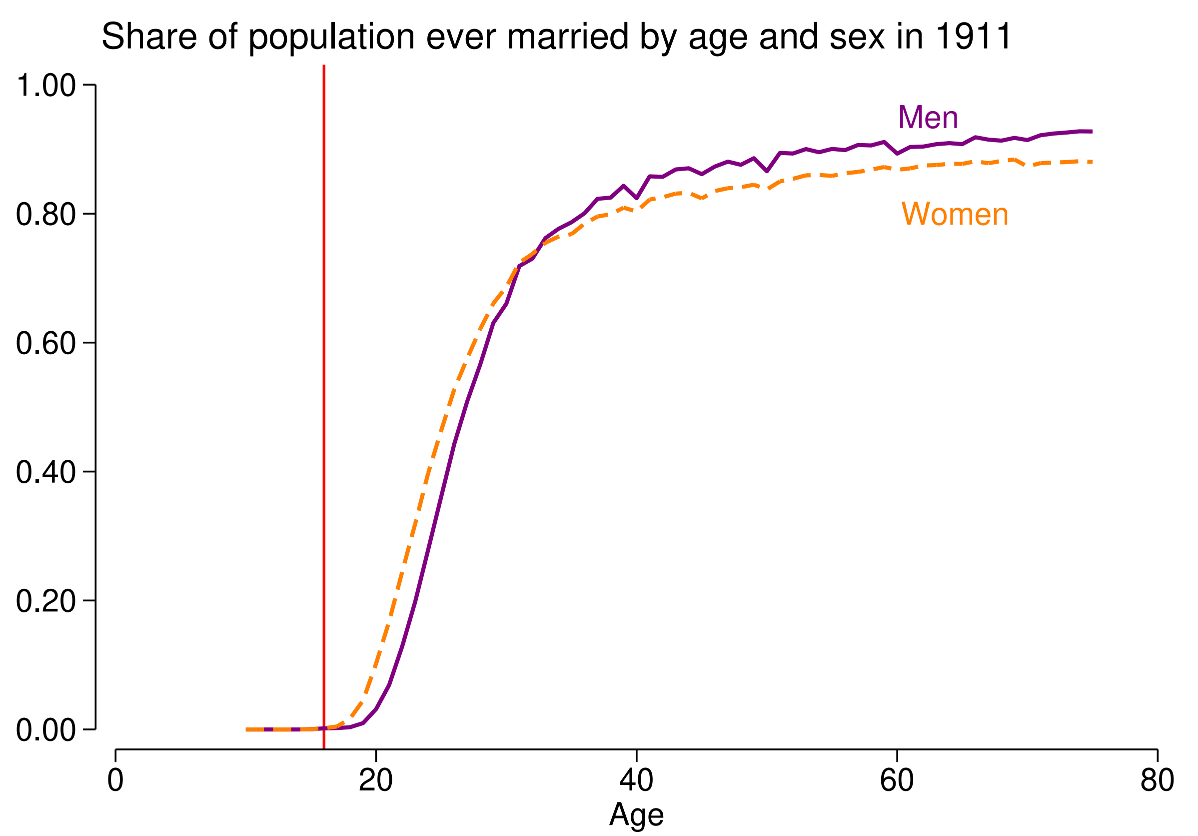
(b) Changing similar names threshold



(c) Changing age-gap threshold for linking

Notes: This figure presents estimated coefficients on the patient indicator variable and 95-percent confidence intervals from robustness specifications for the effects on the likelihood of being enumerated as a single adult. See Table 3 for a description of the empirical specifications and Figures 1 to 3 for a description of the robustness exercises.

Figure A15: Share of population ever married by age and sex in 1911



Notes: This figure plots the share of the population of England enumerated in the 1911 census who report ever being married by age, starting at age 15. Separate plots are shown by gender for men (solid purple line) and women (dashed orange line). Ever married consists of the following codes in response to the question on marital status: married with spouse present in household, married without spouse present, divorced, or widowed. The excluded group consists of individuals who report single marital status. Individuals with missing age, gender, or marital status are excluded from the analysis. The vertical red line indicates age 16 to highlight the restriction that individuals must be no older than age 15 when enumerated in the childhood census in the analysis of marital status.