NBER WORKING PAPER SERIES

CHILDHOOD HEALTH AND SIBLING OUTCOMES: THE SHARED BURDEN AND BENEFIT OF THE 1918 INFLUENZA PANDEMIC

John Parman

Working Paper 19505
http://www.nber.org/papers/w19505

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
October 2013

I have benefited greatly from comments and suggestions from Lucie Schmidt, Trevon Logan, and seminar participants at Virginia Commonwealth University, Dalhousie University, the Washington Area Economic History Seminar, the annual meetings of the Population Association of America and the NBER Summer Institute meetings for the Development of the American Economy program. The views expressed herein are those of the author and do not necessarily reflect the views of the National Bureau of Economic Research.

Childhood Health and Sibling Outcomes: The Shared Burden and Benefit of the 1918 Influenza Pandemic

John Parman

NBER Working Paper No. 19505
October 2013
JEL No. I1,J13,J24,N3,N32

ABSTRACT

There is a growing body of evidence showing that negative childhood health shocks have long term consequences in terms of health, human capital formation and labor market outcomes. However, by altering the relative prices of child quality across siblings, these health shocks can also affect investments in and the outcomes of healthy siblings. This paper uses the 1918 influenza pandemic to test how household resources are reallocated when there is a health shock to one child. Using a new dataset linking census data on childhood households to health and education data from military enlistment records, I show that families with a child in utero during the pandemic shifted resources to older siblings of that child, leading to significantly higher educational attainments and high school graduation rates for these older siblings. There are no significant effects for younger siblings born after the pandemic. These results suggest that the reallocation of household resources in response to a negative childhood health shock tended to reinforce rather than compensate for differences in endowments across children.

John Parman
Department of Economics
P.O. Box 8795
College of William and Mary
Williamsburg, VA 23187
and NBER
jmparman@wm.edu

Childhood Health and Sibling Outcomes: The Shared Burden and Benefit of the 1918 Influenza Pandemic

John Parman*

September 30, 2013

Abstract

There is a growing body of evidence showing that negative childhood health shocks have long term consequences in terms of health, human capital formation and labor market outcomes. However, by altering the relative prices of child quality across siblings, these health shocks can also affect investments in and the outcomes of healthy siblings. This paper uses the 1918 influenza pandemic to test how household resources are reallocated when there is a health shock to one child. Using a new dataset linking census data on childhood households to health and education data from military enlistment records, I show that families with a child in utero during the pandemic shifted resources to older siblings of that child, leading to significantly higher educational attainments and high school graduation rates for these older siblings. There are no significant effects for younger siblings born after the pandemic. These results suggest that the reallocation of household resources in response to a negative childhood health shock tended to reinforce rather than compensate for differences in endowments across children.

1 Introduction

An increasing body of empirical evidence suggests that early childhood health has strong effects on later life outcomes in terms of both socioeconomic status and health. Poor health as a child can translate into lower educational attainment, lower income, higher rates of unemployment and disability, and chronic health problems. The mechanisms underlying these links are a complicated combination of biological effects, with childhood health shocks influencing cognitive development, and the effects of parental investment decisions, with parents changing levels of educational

*Department of Economics, College of William & Mary and NBER, jmparman@wm.edu; This is a working paper, please do not cite without author's permission. For the most recent version of the paper as well as additional figures and information on the data sources, please see http://jmparman.people.wm.edu/. I have benefited greatly from comments and suggestions from Lucie Schmidt, Trevon Logan, and seminar participants at Virginia Commonwealth University, Dalhousie University, the Washington Area Economic History Seminar, the annual meetings of the Population Association of America and the NBER Summer Institute meetings for the Development of the American Economy program.

investment in a child in response to the lower human capital endowment and potentially lower returns on human capital investment in children experiencing bad health shocks. This latter mechanism suggests that the effects of a negative health shock for one child will not be limited to that child alone. If parents are altering their investment decisions it is quite possible that the investments in and outcomes of healthy siblings will also be affected.

The direction of these effects is ambiguous. Parents may substitute investment away from a child in poor health toward healthy siblings given the higher return on investments in those children. However, if parents have an aversion to inequality in outcomes across their children, a poor health shock to one child may lead to a reduction in investments in healthy children relative to investment in the unhealthy child. Understanding investment responses to a negative health shock is therefore an empirical question.

This paper examines that question in the context of the 1918 influenza pandemic, a health shock that had particularly severe effects on children in utero at the time of the pandemic. By linking adult health and educational attainment information from military enlistment records to childhood household information in the 1930 federal census, it is possible to examine how educational investment and health outcomes differ for individuals that had a sibling exposed to the pandemic while in utero and those individuals who did not, offering a way to determine whether parents increased or decreased investment in healthy children in reponse to a negative health shock for one child.

The results suggest that families altered their investments in children, shifting resources toward their older, healthier children. Individuals achieved greater educational attainments if a younger sibling was in utero during the pandemic and experienced small reductions in educational attainment if an older sibling was in utero during the pandemic. The magnitudes of these effects are quite large with older siblings receiving an additional quarter year of education if they had a sibling in utero during the pandemic, a gain in educational attainment similar in magnitude to the decline in educational attainment experienced by individuals in utero during the pandemic. These findings suggest that in the case of the 1918 influenza pandemic, families' responses to a negative health shock to one child served to reinforce differences in endowments across their children; while the pandemic was a burden on children exposed in utero, it was actually a benefit to their older siblings.

Similar patterns did not exist for parents' investment in the health of their children, at least as proxied by adult height. Height does exhibit a similar quantity-quality tradeoff as educational attainment in the data, with both increased family size and lower household incomes being associated with shorter adult heights. However, unlike the results for educational attainment, having a sibling in utero during the pandemic led to no additional effects on adult height. While the reallocation of family resources in response to the pandemic did impact educational outcomes across siblings, it did not impact the net nutrition or health of siblings in a way that translated into differences in attained height.

2 Childhood Health and Adult Outcomes

2.1 Empirical Evidence for the Relationship Between Childhood Health and Adult Socioeconomic Status

There is a growing body of evidence suggesting that poor health early in life, particularly when children are still in utero or infants, can have lasting consquences for health and economic outcomes. Studies have focused on both the initial health endowment of children proxied by such measures as birthweight, Apgar scores and gestational length as well as shocks to childhood health coming from exposure to disease, famine or health interventions to establish the relationship between childhood health and adult outcomes.

Behrman & Rosenzweig (2004) and Royer (2009) use differences in the birthweight of twins to show a positive effect of fetal growth on educational attainment. Black et al. (2007) also use twin data to link differences in birthweight to a broader set of outcomes, demonstrating a positive relationship between birthweight and height, cognitive ability, educational attainment and earnings. Differences in Apgar scores and gestational length between twins and other siblings have also been shown to predict differences in high school completion and social assistance takeup (Oreopoulos et al., 2008).

These studies all establish a strong relationship between initial health endowments of children and adult outcomes. A related literature demonstrates that negative health shocks can be an important source of variation in these health endowments. Work on shocks to childhood nutrition

caused by drought, civil war and price shocks has found that negative shocks to childhood health lower school enrollment rates and lead to lower levels of completed schooling (Alderman et al., 2001, 2006). Case et al. (2005) find adverse effects of chronic illness as a child on educational attainment and socioeconomic status of individuals as adults. Public health interventions that reduce the frequency of negative childhood health shocks have also been shown to improve educational outcomes (see for example Miguel & Kremer (2004) examining the introduction of deworming drugs in Kenyan schools and, closer to the population of interest in this study, the work on hookworm and malaria eradication in the South in the early twentieth century by Bleakley (2007, 2010)).

2.2 The Mechanisms Linking Childhood Health to Adult Outcomes

There are a variety of mechanisms that may be responsible for these observed links between poor health endowments or negative childhood health shocks and adult outcomes. A thorough overview of the possible mechanisms and the empirical support for each can be found in Currie (2009). The most obvious mechanism relates to the direct link between childhood health and future health outcomes and the economic outcomes influenced by those health outcomes. The fetal origins literature suggests that poor in utero health conditions can lead to chronic health problems including obesity, heart disease and diabetes (see for example Barker (1998a) and Barker (1998b)). These health problems could have a direct impact on productivity and earnings. There is also evidence that poor in utero and childhood health can cause cognitive impairment such as the study of Romanian orphans by OConner et al. (2000) finding lower cognitive scores and general developmental impairment among adoptees who faced longer periods of poor childhood health.[1]

However, it is not solely through persistent health problems and developmental impairment that childhood health can influence adult outcomes. Poor childhood health can alter the human capital investment decisions made by parents. Thus even if a childhood health shock is temporary, it may still have lasting effects through lower parental investments in human capital, particularly

[1]For an overview of work on the fetal origins hypothesis in both the epidemiology and economics literatures, see Almond & Currie (2011).

in terms of formal schooling. There is a large theoretical literature on the economics of the family extending back to the work of Becker and Tomes in which human capital investment decisions are dependent on the health endowment of children and the marginal costs of increasing a child's human capital stock which may itself be a function of a child's health (Becker & Tomes, 1976; Becker, 1991).

In the model developed by Becker and Tomes, the impact of a low health endowment could potentially increase parental investments in that child relative to his siblings. The basic logic is that parents' utility is an increasing, concave function with respect to each child's quality. Thus the return on investment in a child's quality will initially be relatively high for a child with a low health endowment. If the parents are going to equalize the marginal return on investment across all children in order to maximize utility, they will investment more in the low endowment children and relatively less in high endowment children.

This becomes more complicated when the return on human capital investment is a function of a child's health endowment. If healthier children experience greater returns from human capital investment, perhaps by being able to focus better in classes or having better school attendance, then it will require greater human capital investment in high endowment children relative to low endowment children to equate the marginal return on investment across children and maximize household utility. In this case, parents' human capital invesment decisions would serve to reinforce health disparities across children. A child receiving a negative health shock would receive less educational investment than his siblings.

Subsequent work on modeling household investment decisions has incorporated parents' aversion to inequality in outcomes across their children. Examples include Behrman et al. (1982), Ejrnæs & Pörtner (2004), and Adhvaryu & Nyshadham (2011). These models allow for parents to value equality of outcomes across children, leading to the possibility that even if poor childhood health effectively increases the price of a unit of human capital, parents may choose to invest more in their less healthy children. The work of Ejrnæs & Pörtner is particularly interesting in the context of this paper as their model includes decisions over family size. The degree of inequality aversion on the part of the parents influences whether they decide to have another child after having a child with a particularly high or low health endowment. Ejrnæs & Pörtner's model therefore raises the possibility that it is not just the distribution of resources

across children but also the number of children that can be affected by a negative health shock for one child. The health shock may therefore have different impacts on children born before or after the afflicted child. As Section 6 will show, this does turn out to be the case for the 1918 influenza pandemic.

Beyond helping to explain the ways that a child's health can impact long run outcomes through parents' investment decisions, these works on the allocation of household resources also suggest that siblings will be affected by a child's health shock as parents either substitute resources away from these siblings to achieve greater equality of outcomes across children or invest more in these children because of their relatively higher rate of return on human capital investments. It is therefore possible that the human capital investments in a child and the labor market outcomes dependent on those investments will be significantly impacted by a health shock to a sibling.

There is far less empirical evidence on the consequences of health shocks on investments in healthy siblings. A major obstacle has been appropriate data. The majority of studies assessing the impact of negative health shocks on educational investments and subsequent labor market outcomes focus on sibling or twin data in order to control for household fixed effects. The drawback is that twin data allows for identifying the difference in investment between siblings when a health shock occurs but not a change in the overall level of investment across all siblings. The alternative is to focus on a health shock that is easily observed and affects a single child in the household. Adhvaryu & Nyshadham (2011) take this approach, using Tanzania's distribution of iodine substitutes to pregnant mothers as a positive shock to fetal health. Adhvaryu & Nyshadham find that parents of children that received this positive health shock made greater health investments in both the treated children and their siblings. This paper will follow a similar approach to identify how educational investments across siblings responded to the influenza pandemic of 1918.

3 Childhood Health and the Influenza Pandemic of 1918

A major limitation to studying changes in household resource allocation across children is finding a health shock that is both exogenous and focused on a single child. The influenza pandemic of 1918 provides one such shock. Almond (2006) and Almond & Mazumder (2005) have used the influenza pandemic as a test of the fetal origins hypothesis, demonstrating that children exposed to influenza while in utero suffered from significant long term effects. Children in utero during the pandemic had lower educational attainments, higher rates of physical disability and lower socioeconomic status relative to cohorts born before or after the pandemic. While the analysis of Almond (2006) and Almond & Mazumder (2005) focused on the American experience, similar effects have been demonstrated for Europe and South America (Neelsen & Stratmann, 2011; Nelson, 2010).

The influenza pandemic of 1918 was remarkable for its severity both in terms of the number of infected people and the mortality rate for those infected. Worldwide the pandemic claimed approximately 50 million lives (Johnson & Mueller, 2002). In the United States, over 25 percent of the population contracted the virus and over 600,000 Americans died. The virus spread through the United States rapidly. After a mild wave of influenza in the late spring of 1918, a second deadly wave began in September of 1918 in Boston and spread throughout the entire country within a month. As shown in Figure 1, deaths from influenza rose dramatically from September of 1918 to October of 1918, remaining high in November and then falling in the subsequent months as the second wave dissipated.

A unique feature of the pandemic was the variation in infection and mortality rates by age. While deaths from influenza are typically concentrated among infants and the elderly, the Spanish flu was unique in that it targeted individuals in their twenties and thirties. Figure 2 depicts the unusual W-shape of deaths from influenza during the pandemic compared to the traditional U-shape. The pandemic targeted individuals of childbearing age. Beyond targeting this group as a whole, pregnant women particularly suffered. In studies of hospitalized pregnant women during the pandemic, death rates ranged from 23 to 71 percent and of the surviving women, 26 percent lost their child (Barry (2005), p. 240). These statistics suggest that children in utero during the

pandemic received a severe negative health shock due to the dramatic effects of the pandemic on maternal health.[2]

The rapid and unexpected onset of the 1918 influenza pandemic, its severe effects on pregnant women and the relatively short duration of the pandemic make the pandemic a particularly useful health shock for studying the effects of poor childhood health on adult outcomes. The work of Almond, Almond & Mazumder, Neelsen & Stratmann and Nelson has convincingly shown that the health effects on children in utero during the pandemic translated into significantly worse health, educational and socioeconomic outcomes as adults for those children relative to cohorts born just before or after the pandemic. These findings suggest that the pandemic would be an equally useful setting for identifying the effects of a negative health shock on siblings given its dramatic health effects targeted at a single child within a family.

4 Data Set Construction

4.1 Data Sources

While the influenza pandemic of 1918 provides an interesting case of a major health shock targeted at a very specific birth cohort, it presents several unique issues in terms of obtaining relevant data. The studies of Almond and Mazumder utilize modern records to estimate the educational and socioeconomic outcomes of individuals in utero during the pandemic. They rely on the availability of federal census data from the 1960s, 1970s and 1980s containing educational attainment, income and disability information as well as data on health outcomes from Survey of Income and Program Participation data. In these modern data sets it is possible to identify those individuals born during the pandemic and to compare their health and socioeconomic status as adults to individuals born before or after the pandemic. While these datasets provide a wealth of interesting adult outcomes, they do not offer any information about siblings or childhood household characteristics, information that is crucial to the assessing how the pandemic changed resource allocations across children.

A solution to this problem is to create a panel of historical data in which individuals are

[2]For a much more thorough discussion of the features of the pandemic that make it a plausibly exogenous and unanticipated shock to fetal health, see Almond (2006).

observed both as children in their parents' household and as adults. The 1930 federal census offers an opportunity to observe individuals born in the years surrounding the pandemic as adolescents still living with their parents. From the original census manuscripts it is possible to identify the ages and genders of all of an individual's siblings, information on parents including occupations, ages and literacy, as well as basic household information such as the location of the household and the value of the house.

Measuring adult outcomes requires an alternative data source. There are few historical data sources from this period that contain the sort of education, income and health information that would be of interest. The most detailed individual records come from the federal census but the federal census did not start asking for educational attainment or income until 1940 and did not record relevant health information until 1970 with the inclusion of disability questions. With the 72 year rule, it is not possible to match individual children from the 1930 census to the more recent censuses utilized by Almond containing both adult educational attainment and health data.[3] An alternative is to use military enlistment records from World War II, a period in which the individuals born in the years surrounding the 1918 pandemic were young adults.[4]

The National Archives and Records Administration has digitized the information from the enlistment cards used by the United States Army during World War II. One of the main purposes of the enlistment cards was for the Adjutant General's Office to create "tables analyzing occurrence of the various characteristics among individuals enlisted or inducted, and to provide data for policies of demobilization" (U.S. War Department, p. 12). Consequently they contained information on the education, occupations and health of enlistees relevant to assessing the strength of the Army and efficiently assigning enlistees to various positions in the army. The years of secondary and postsecondary education, civilian occupation, height and weight of each enlistee were recorded along with basic demographic information such as year of birth, state of birth and state of residence.[5] With 65 percent of the males born in the five years surrounding

[3]The 72 year rule states that the federal government will not release personally identifiable information about an individual until 72 years after it is collected for the decennial census. This information is necessary to link individuals from their childhood households to their adult census records.

[4]While these enlistment records are from roughly the same time period as the 1940 federal census and contain similar educational attainment data, the extra two to three years between 1940 and the typical year of enlistment actually substantially increases the number of individuals born in the years around the 1918 influenza pandemic who can be observed with completed educational careers.

[5]To assess whether the education and stature data provided in the enlistments records will show similar responses to the pandemic as the census information recorded decades later, regressions comparable to those in Almond (2006) were

the influenza pandemic serving in the military during World War II, these enlistment records provide a tremendous data source for measuring adult outcomes for the population of interest.[6]

4.2 Matching Enlistee Records to the Federal Census

The process of matching enlistment records to census records begins by sampling the enlistment records. First, records missing key information for the matching procedure or for the analysis are dropped. This includes dropping any observations for which name, year of birth, state of birth, educational attainment, or height and weight are missing. A one percent sample of the remaining observations is taken by sorting the observations alphabetically by last name and sampling every hundreth record.

Each individual in this sample is then searched for in the 1930 federal census. The geneology website ancestry.com has created an electronic index of every individual in the 1930 census. This index can searched by name, birth year, birth state, race and gender, all of which are available from the enlistment records. If a unique match is found information on the individual's family is is recorded. For the first 25,000 enlistees, this matching process was done by hand. This sample of hand matched enlistees was used to estimate the parameters of a matching algorithm to automate the process of identifying matches and determining whether a match is unique. This matching algorithm was then used to match the remainder of the sample. Details on the construction of the matching algorithm and summary statistics on the reliability of the algorithm are provided in the appendix.

Once the enlistees have been matched to the federal census several pieces of information are recorded from the census. The household location and the names, ages and genders of the individuals' siblings have all been transcribed by ancestry.com and can be pulled from the search results using a computer script. Parents' occupations and literacy, whether the family rents or

run using the Army enlistee data to test for the effect of being in utero during the pandemic on educational attainment and stature. Regression results are presented in Table 1. The results for educational attainment are nearly identical to those found by Almond, with being in utero during the pandemic being associated with 0.13 fewer years of schooling and a reduction in the probability of completing high school of 2.8 percent. The enlistment records also show that being in utero during the pandemic impacted height and weight. Individuals in the influenza cohort were significantly shorter than individuals born earlier or later, had lower weights and had lower body mass index values.

[6]The 65 percent figure is based on tabulations from using the the World War II veteran status question from the 1950 census including only individuals that answered either 'yes' or 'no' (rather than 'N/A'). The tabulations are calculated using the 1% census sample from the Integrated Public Use Microdata Series.

owns their home, and house value are transcribed by hand from an image of the original census manuscript page.

4.3 Sample Selection Bias

A major concern with this approach of linking individuals across historical data sources is sample selection bias resulting from the probability of being successfully linked being correlated with individual characteristics. Linking enlistees to the federal census can fail for a variety of reasons. The simplest reasons for not finding a unique match are illegible census records or finding multiple individuals that match an enlistee's personal details. Neither of these reasons generates major concerns in terms of the representativeness of the matched sample. Legibility of the census records is a function of the enumerator's handwriting and the image quality of the scans of census pages, neither of which are related to the characteristics of the individual being matched. Multiple census matches is a product of having a common name. One possible concern with this is that people from states with larger populations will be less likely to be matched. This is easily accounted for by controlling for state in the analysis.

The more troubling reasons for failing to match an individual relate to the accuracy of the information they report. Particularly problematic is age misreporting. If parents misreport age to the census enumerator or the individual misreports his age when enlisting, it will not be possible to match the enlistment records to the census record. To the extent that misreporting age is correlated with educational attainment or unobserved dimensions of ability, the resulting sample will be underrepresentative of low ability individuals. These aspects of ability will certainly influence household resource allocation decisions. Furthermore, if the effects of health shocks on educational investments are concentrated in the lower tail of the educational distribution, there is a concern that the affected individuals will not make it into the matched sample.

To assess the extent to which the matching process affects the composition of the sample, Table 2 presents summary statistics for individuals who were successfully matched to the 1930 federal census and individuals who could not be matched. 48 percent of the enlistees could be matched to the federal census, a match rate that compares favorably with other studies using similar census linking techniques. In terms of age, year of enlistment and physical characteristics the matched individuals are indistinguishable from the unmatched individuals. Matched indi-

viduals are slightly more educated on average and slightly less geographically mobile than the unmatched enlistees. A more detailed depiction of the differences in educational attainment is given by Figure 3 showing the distribution of educational attainment for enlistees by match status. Matched individuals are more likely to attend high school and more likely to be high school graduates. However, the matched enlistees still cover the same overall range of educational attainments as the unmatched enlistees. Even if effects of health shocks are focused on individuals in the tails of the educational attainment distribution, these effects should be observable in the matched sample of enlistees.

A separate concern with the sample is the reliance on military enlistment records for educational attainment and adult health outcomes. Military enlistees are not a random sample of the population as a whole. An individual had to meet minimum physical requirements to enlist in the military. Individuals could be rejected on the basis of height, weight, defective teeth, poor vision, deafness, venereal disease and other conditions. Enlistees could also be rejected for being illiterate. The set of enlistees will therefore be underrepresentative of the least healthy individuals and potentially least educated individuals in the population, a group that may be the most impacted by negative childhood health shocks. While this would be quite problematic when focusing on the effects of a health shock on an individual's own outcomes, it is less of a concern here because the enlistment records are being used to examine the siblings of individuals in utero during the influenza pandemic.

5 Estimation Strategy

The basic empirical approach will be to estimate adult outcomes as a function of individual and household characterstics and an indicator variable for whether the individual had a sibling in utero during the influenza pandemic. The basic regression equation would therefore be

$$Y_{i,j,k} = X_i'\beta + Z_j'\gamma + \alpha Flu_j + \mu_k + \varepsilon_{i,j,k} \tag{1}$$

where $Y_{i,j,k}$ is the outcome of interest for individual i from household j living in state k. The main outcome of interest will be educational attainment but the height and weight data will also be used as proxies for adult health outcomes. X_i is a vector of individual characteristics which,

depending upon the particular specification used, includes a polynomial in birth year, race, birth order and birth order among brothers. Available information for the vector of household characteristics, Z_j, includes parental incomes, house value and family size.[7] State fixed effects, μ_k, are included to account for variation in average health and educational attainment across states. Flu_j is a dummy variable indicating whether there was a child in household j in utero during the influenza pandemic.

It is the coefficient on Flu_j that will pick up the effects of a negative health shock on resource allocation across children. A positive estimate for α would indicate that parents are substituting resources away from a child receiving a negative health shock toward healthy children. In this case, investment decisions would serve to reinforce differences in endowments across children. However, if α has a negative sign, it is an indication that parents reduce investments in healthy children in response to a negative health shock for one child, suggesting that parents have an aversion to inequality in outcomes across their children or that the health shock reduces family resources in a way that leads to lower investments across all children.

One issue with the estimation of Equation 1 is censoring of the dependent variable. In the case of the main variable of interest, years of educational attainment, the variable is censored both from below and from above. Educational attainment in the Army enlistment records is reported as years of secondary and postsecondary education up to eight years total. Individuals with more than eight years of secondary and postsecondary education are coded as a nine in the enlistment records. Assuming an individual enters high school after eight years of primary schooling, the observed educational attainment E_i is a censored version of the total number of years of education, E_i^* as follows:

$$E_i = \begin{cases} 0 & \text{if } E_i^* \leq 8 \\ E_i^* - 8 & \text{if } 8 < E_i^* \leq 16 \\ 9 & \text{if } 16 < E_i^* \end{cases} \tag{2}$$

This leads to a substantial number of observations censored from below; over 20 percent of the

[7]The availability of these household characteristics in the linked sample are particularly important given the findings of Brown (2011) suggesting that families having children during the influenza pandemic were actually not representative of the general population. In particular, these families tended to be less prosperous relative to the general population. Controlling for household characteristics will help ensure that any estimated effects of the pandemic are being driven by the impacts on children's health and not by other omitted variables.

enlistees have a zero reported for educational attainment. There is also censoring at the other end of the educational attainment distribution although this censoring is far less problematic as there are only 22 individuals in the matched enlistee sample with more than eight years of secondary and postsecondary education. To account for censoring, tobit models will be used to obtain consistent estimates of α in Equation 1.[8]

A second problem with the estimation of Equation 1 relates to identifying siblings in utero during the influenza pandemic. The severe effects of the pandemic were felt during October, November and December of 1918. Children born any time between October of 1918 and August of 1919 were therefore potentially affected by the pandemic while in utero. The children in utero during the entire pandemic would have been born between January and June of 1919. With modern census data, Almond was able to identify individuals born during the first two quarters of 1919, separating these individuals with high exposure in utero from other children born in 1919 receiving little to no in utero exposure to influenza. The 1930 census does not provide quarter of birth or even year of birth. Instead, the 1930 census reports the age of each child as of April 1, 1930. This makes it difficult to identify exactly who was exposed to influenza while in utero. Children exposed to influenza during their second and third trimesters will have a reported age of 11 in the 1930 census. Children exposed during their first trimester will have a reported age of 10 in the census. **??** shows how in utero exposure varied by date of birth. The graph shows the average monthly mortality rate for the population during the period that the child was in utero and the relative contributions of the mortality rates during the first, second and third trimesters to the overall average (these mortality rates are shown in Figure 1). Children born to the left of the dashed line will have an age of 11 in the 1930 census while children born to the right of the line will have an age of 10 in the federal census.

One possibility is to define Flu_j as equal to one if a sibling has an age of either 10 or 11 in the

[8]Note that there is a second censoring issue related to whether the enlistee has completed his educational career. Given that many of the enlistees are in their late teens and early twenties, it is likely that many had their educational careers interrupted by the war. For these enlistees, the observed educational attainment should be considered an underestimate of their final educational attainment (or at least the family's desired overall level of educational attainment for the enlistee). Estimates have been run for both the complete sample of enlistees and for a restricted sample containing only those enlistees who have completed their educational careers. An enlistee is considered to have completed his educational career if the enlistee has been out of school for two or more years assuming that secondary school was started at the age of 14 and all years of schooling were completed consecutively. Both regression samples yield similar results. The results presented in the paper are based on the sample restricted to individuals with completed careers. Results using the unrestricted sample are quite similar and are available from the author.

1930 census. This would ensure that all children with a sibling in utero during the pandemic are counted in the treatment group for the regressions. However, it would also count a large number of siblings as treated even if they were born before the pandemic began or conceived after the pandemic ended. Misclassifying these children as treated will introduce substantial measurement error in Flu_j and produce a potentially large attenuation bias for the estimated coefficient α.

An alternative is to treat only those siblings with a reported age of 11 as receiving in utero exposure to influenza. While this excludes those siblings exposed only during the first trimester, it has the advantage that even those children incorrectly assigned to the treatment group would have been exposed to the pandemic as infants (unlike the 10-year-olds conceived after the pandemic). These 11-year-olds in the 1930 census, while born before the pandemic began, would still have potentially received a more severe health shock from the pandemic than their older siblings. As Figure 2 shows, infants as a group did suffer severely from the pandemic. This presents a smaller measurement error problem than using the combined 10- and 11-year-old group and for that reason will be the preferred approach for defining Flu_j.

6 Results

6.1 Main Results

To estimate Equation 1, the sample of successfully matched enlistees is restricted to males born between the years of 1912 and 1924. The gender restriction is due to having an extremely small number of female enlistment records that could be matched to the federal census (and that small sample being less representative of the general population). The birth year restriction is based on the number of enlistees in each birth cohort. These years each have over 100,000 records. The number of records falls of significantly for the years before and after this range, leading to concerns that the enlistees would be less representative of the population as a whole as well as far less likely to be affected by an unhealthy sibling born in 1919.

The regression sample is further restricted to individuals who are listed as being the son of the head of household in 1930, a restriction that eliminates roughly ten percent of the matched enlistees. Restricting the sample to sons of heads of household simplifies the types of household resource allocation being studied. The estimates in this section will focus on households in which

the household head is the biological parent of both the individual in the matched enlistee sample and any potential siblings born during the influenza pandemic. Those individuals dropped from the sample are coming from households in which the household head is either a grandparent or another relative of some sort or from households that are not traditional family units but rather boarding houses, prisons or other institutions. These situations would correspond to very different resource allocation problems and are not the focus of this study.

Table 3 presents summary statistics for the enlistees included in the regression sample. The birth years of the enlistees are centered around the birth cohort suffering from the influenza pandemic. The enlistees have on average three siblings. Roughly 15 percent of enlistees had a sibling exposed to the influenza pandemic while in utero or as an infant. For slightly over half of those individuals, it was an older sibling born during the pandemic. To allow for the possibility that parents alter their investment decisions for younger children in a different manner than for older children, the regressions will include separate indictors for having a younger sibling born during the pandemic and for having an older sibling born during the pandemic.

The results of estimating Equation 1 using years of secondary and postsecondary education as the dependent variable are given in Table 4. Columns (1) and (2) restrict the sample of enlistees to those not in utero during the pandemic and therefore not subject to the health shock. Columns (3) and (4) include these enlistees in utero during the pandemic with an additional indicator variable to account for any direct effects of being in utero during the pandemic on educational outcomes. Columns (1) and (3) treat all siblings identically while columns (2) and (4) allow the affect of a sibling to vary with the gender of that sibling.

The marginal effect of having a sibling born during the pandemic is notable for its magnitude but also how it differs depending on whether it was a younger or older sibling. The effect of having an additional sibling in general is large and negative whether that additional sibling is younger or older. This is consistent with the existing literature on quantity-quality tradeoffs for children. What is suprising is that the effect of having a sibling born during the pandemic has a different sign depending whether the sibling is younger or older. While a negative health shock to an older sibling leads to a statistically insignificant reduction in educational investments, a similar shock to a younger sibling, whether that younger sibling is male or female, actually leads to a large increase in educational attainment.

The magnitudes of these effects are substantial. The increase in educational attainment associated with having a younger sibling born during the pandemic is over a quarter of a year. This is actually a larger effect than the marginal effect of being born during the pandemic on one's own educational attainment as estimated by Almond (2006) with modern census data and replicated using the set of all enlistee records in Table 1.[9] These results underscore how significant the reallocation of household resources is when a child receives a negative health shock. Focusing solely on the outcomes of the unhealthy child ignores equally large effects experienced by that child's siblings.

Given the substantial number of enlistees with either zero years of secondary education or exactly four years of high school, it seems reasonable to also consider treating educational attainment as a choice between attending high school or not or a choice between completing high school or not. With this is mind, an alternative set of regressions is estimated using binary outcomes as the dependent variable rather than years of education. Table 5 presents estimates of a logit model in which the dependent variable is an indicator for whether the enlistee graduated high school. Table 6 presents similar estimates using an indicator for whether the enlistee attended high school as the dependent variable. Marginal effects computed using these coefficients are presented in Table 7. Both of these sets of regressions yield similar results to the tobit estimates. In both cases, having a greater number of older or younger siblings significantly decreases the likelihood of either attending or graduating from high school. The same is true of having an older sibling born during the pandemic although these coefficients are not statistically significant. However, if an enlistee has a younger brother or sister in utero during the pandemic, the enlistee is more likely to attend and graduate from high school.[10] The magnitude of these effects is once again quite large. Having a younger sibling in utero during the pandemic was associated with a six percent increase in both the probability of attending and the probability of graduating

[9]Note that the impacts in Table 1 may underestimate the impact of the pandemic due to sample selection. For the more modern census data, the least healthy individuals impacted by the pandemic may have died. For the enlistee records, those least healthy individuals may have been rejected by the military. In both cases, this would lead to an underestimate of the negative impact of being in utero during the pandemic as the regression sample would be composed of people healthier on average than the true population of treated individuals.

[10]It is worth noting that years of education and high school attendance both respond similarly to having a younger sister or brother in utero during the pandemic. However, the impact of a younger sister in utero during the pandemic is noticeably larger in magnitude and more statistically sigificant than the impact of a younger brother in utero during the pandemic on the likelihood of high school graduation. While the data do not have sufficient detail to unravel the reasons behind this difference, it suggests that families' responses to health shocks are dependent on the gender of the child suffering the health shock.

from high school. This reinforces the conclusion that families were shifting educational investment toward older children in response to a negative health shock to one child. These effects are substantial: the marginal effect of having a younger sibling in utero during the pandemic is roughly equal to the marginal effect of a twenty percent increase the household head's income on attending on graduating high school.

While limited in scope, there is basic adult health information in the enlistment records in the form of height and weight that can be used to test for similar effects of the reallocation of household resources on the health outcomes of siblings when one child receives a health shock. As a product of net nutrition over the course of childhood, adult height can serve as a proxy of childhood health conditions potentially influenced by parents' resource allocation decisions. The weight information in the enlistment records also allows for measuring body mass index, offering a measure of health that is influenced by adult health behaviors and less sensitive to parental investment decisions.[11]

The results of estimating Equation 1 using these health measures as the dependent variables are given in Table 8 and Table 9. The height regressions suggest quantity-quality tradeoffs similar to those found for educational attainment: an additional sibling whether younger or older is associated with a substantial reduction in height suggesting that health investments in each child decline as family size increases. Also consistent with the education estimates, an increase in household head's income is associated with a significant increase in adult height. These large and significant coefficients suggest that height may be able to pick up differences in health investments across siblings. However, there are no statistically significant effects of having a sibling in utero during the pandemic on an enlistee's own height. The body mass index results do not produce any coefficients on the sibling variables that are either large in magnitude or statistically significant. This is unsurprising given that body mass index, unlike height, is as much a product adult health behaviors as childhood health investments. These results for the basic health measures suggest that health investements across children were far less sensitive to

[11] As identified by Ferrie et al. (2011), the weight field on the enlistment cards was actually used to record scores on the Army General Classification Test (AGCT), a test of general mental acuity, for a brief period of time between March and June of 1943. Individuals enlisting during this period are therefore dropped from the regressions using weight or body mass index as a dependent variable. The AGCT scores themselves would be an interesting outcome but the short time span over which they were recorded leaves an insufficient number of observations to work with. There are only 137 individuals in the linked sample that have AGCT scores rather than weight information.

a childhood health shock than educational investments.

6.2 Robustness Checks

The results from the educational attainment regressions present strong evidence that reallocation of educational resources was taking place in response to a health shock to a child. However, there are potential concerns about whether this is a proper way to interpret the coefficient on the flu sibling variable. A primary concern is that although a substantial number of people including pregnant mothers were infected with influenza in 1918, the majority of the population was not infected meaning that many of those siblings being counted as part of the influenza cohort did not actually receive a negative health shock. The flu sibling variable is therefore measured with error and the coefficient on the variable will suffer from an attenuation bias, suggesting that if anything the marginal effect of a sibling's negative health shock on an individual's own education may be substantially larger than the point estimates in the previous section.

A more troubling possibility is that the coefficients on the flu sibling variable are picking up the effects of something other than having a sibling in poor health. In particular, for those individuals with a younger sibling in utero during the pandemic, the variable may be picking up an effect related to being of high school age at the start of the Great Depression while for those individuals born after the pandemic the variable would be picking up effects related to being of high school age at the end of the Great Depression. Similarly, one might worry that the coefficients are picking up differential effects of World War I or the major changes in American educational institutions taking place during this time period. The possibility that the younger versus older sibling coefficients are picking up a dimension of changing conditions across cohorts rather than a single health shock require careful consideration. Two different sets of estimates are provided in the appendix to attempt to rule out these types of concerns.

The first approach is to more flexibly control for year of birth. All of the regressions are rerun including birth year fixed effects rather than simply a cubic in birth year (results are given in Table 14, Table 15 and Table 16 of the appendix). These birth year fixed effects will absorb any systematic differences across birth cohorts related to differences in economic conditions or educational institutions across different cohorts. The inclusion of these birth cohort fixed effects does not change any of the findings in any substantial way. The coefficients on the younger flu

sibling variables remain positive, large and statistically significant across all of the specifications.

A second approach involves constructing placebo treatment groups by redefining the birth year for the flu cohort. By shifting the birth year for the flu cohort it is possible to test whether the flu cohort variable is picking up effects specific to those individuals in utero during the pandemic or whether the older and younger flu sibling coefficients are the product of a more mechanical relationship that would arise for any specific sibling cohort included in the regressions. In the appendix, all of the regressions are rerun shifting the birth year for the flu cohort by two years in either direction.[12] The coefficients on the sibling in utero during the pandemic variables are summarized in Table 10 with complete regression results provided in the appendix. When shifting the influenza year either ahead or back by two years, the large effects of having a younger sibling in utero during the pandemic completely disappear. The coefficient magnitudes drop by an order of magnitude and become statistically indistinguishable from zero.[13] These robustness checks suggest that the results are not being driven by general time trends or a mechanical relationship arising from controlling for a specific sibling birth cohort, bolstering the conclusion that the earlier estimates are capturing the effects of an unhealthy sibling.

7 Conclusions

The influenza pandemic of 1918 provides a unique opportunity to examine the effects of a negative health shock for one child on the outcomes of his or her siblings. The pandemic was an enormous, unanticipated health shock that was particularly severe for women of childbearing age leading to substantially compromised health for children in utero during the pandemic. The results from this newly linked dataset suggest that parents altered their investments across all children in response to that shock.

The investments were not altered in the same way for all children. The positive effect on educational attainment of having a younger sibling born during the pandemic and the negative

[12]Two years is used rather than one to ensure that there are no actual influenza children in the placebo group. Both the fact that the influenza exposure does not match up perfectly with the 1919 birth year and the possibility of age misreporting in the census could lead to a one-year shift leaving many children actually exposed to influenza in utero in the placebo influenza cohort.

[13]There is one regression in which marginal effect of having an older sibling in the placebo cohort is statistically significant (Table 17, Column 6). However, one coefficient significant at the five percent level out of the 24 placebo influenza cohort coefficients can most likely be attributed to Type I error.

effect on educational attainment of having an older sibling born during the pandemic suggest that parents were substituting resources for educational investment away from younger children toward their older children in response to the health shock. The results for the effects on height do not reveal a similar reallocation across children in terms of investments in health though this may be a product of height being insufficiently sensitive to changes in health investments.

The magnitudes of these effects are quite large and of the same order of magnitude as the effects of being in utero during the pandemic on an individual's own educational outcomes. These findings highlight the importance of considering the effects of early childhood health interventions on not just the treated child but on siblings as well. The experience of the influenza pandemic reveals that a health shock to one child can have major consequences for the distribution of resources across other siblings and the overall equality of outcomes within a family. Families' responses to the influenza pandemic served to reinforce differences in endowments across children, turning the burden of disease for one child into a benefit for his older siblings.

References

Adhvaryu, A., & Nyshadham, A. (2011). Endowments and investment within the household: Evidence from iodine supplementation in tanzania. *Working Papers*.

Alderman, H., Behrman, J., Lavy, V., & Menon, R. (2001). Child health and school enrollment: A longitudinal analysis. *Journal of Human Resources*, (pp. 185–205).

Alderman, H., Hoddinott, J., & Kinsey, B. (2006). Long term consequences of early childhood malnutrition. *Oxford Economic Papers*, *58*(3), 450–474.

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

Almond, D., & Currie, J. (2011). Killing me softly: The fetal origins hypothesis. *The Journal of Economic Perspectives*, *25*(3), 153–172.

Almond, D., & Mazumder, B. (2005). The 1918 influenza pandemic and subsequent health outcomes: an analysis of sipp data. *American Economic Review*, (pp. 258–262).

Barker, D. (1998a). In utero programming of chronic disease. *Clinical science*, *95*(2), 115–128.

Barker, D. (1998b). *Mothers, babies, and health in later life*. Elsevier Health Sciences.

Barry, J. (2005). *The great influenza: the epic story of the deadliest plague in history*. Penguin Group USA.

Becker, G. (1991). *A Treatise on the Family*. Harvard Univ Pr.

Becker, G., & Tomes, N. (1976). Child endowments and the quantity and quality of children. *The Journal of Political Economy*, *84*(4), 143–162.

Behrman, J., Pollak, R., & Taubman, P. (1982). Parental preferences and provision for progeny. *The Journal of Political Economy*, (pp. 52–73).

Behrman, J., & Rosenzweig, M. (2004). Returns to birthweight. *Review of Economics and Statistics*, *86*(2), 586–601.

Black, S., Devereux, P., & Salvanes, K. (2007). 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.

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

Brown, R. (2011). The 1918 us influenza pandemic as natural experiment, revisited. *Mimeo, Duke University*.

Case, A., Fertig, A., & Paxson, C. (2005). The lasting impact of childhood health and circumstance. *Journal of Health Economics*, *24*(2), 365–389.

Currie, J. (2009). Healthy, wealthy, and wise: Socioeconomic status, poor health in childhood, and human capital development. *Journal of Economic Literature*, *47*(1), 87–122.

Ejrnæs, M., & Pörtner, C. (2004). Birth order and the intrahousehold allocation of time and education. *Review of Economics and Statistics*, *86*(4), 1008–1019.

Ferrie, J. P., Rolf, K., & Troesken, W. (2011). Cognitive disparities, lead plumbing, and water chemistry: Intelligence test scores and exposure to water-borne lead among world war two us army enlistees. *National Bureau of Economic Research WP No. 17161*.

Johnson, N., & Mueller, J. (2002). Updating the accounts: global mortality of the 1918-1920" spanish" influenza pandemic. *Bulletin of the History of Medicine*, *76*(1), 105–115.

Lait, A., & Randell, B. (1996). An assessment of name matching algorithms. *Technical Report Series-University of Newcastle Upon Tyne Computing Science*.

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

Neelsen, S., & Stratmann, T. (2011). Long-run effects of fetal influenza exposure: Evidence from switzerland. *Social Science & Medicine*.

Nelson, R. (2010). Testing the fetal origins hypothesis in a developing country: evidence from the 1918 influenza pandemic. *Health economics*, *19*(10), 1181–1192.

OConner, T., Rutter, M., Beckett, C., Keaveney, L., & Kreppner, J. (2000). The effects of global severe privation on cognitive competence: extension and longitudinal follow-up. *Child Dev*, *71*(2), 376–90.

Oreopoulos, P., Stabile, M., Walld, R., & Roos, L. (2008). Short-, Medium-, and Long-Term Consequences of Poor Infant Health: An Analysis Using Siblings and Twins. *Journal of Human Resources*, *43*(1), 88.

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

Ruggles, S., Sobek, M., Alexander, T., Fitch, C., Goeken, R., Hall, P., King, M., & Ronnander, C. (2009a). Integrated public use microdata series sample of the 1930 federal census. Accessed throgh usa.ipums.org/usa/.

Ruggles, S., Sobek, M., Alexander, T., Fitch, C., Goeken, R., Hall, P., King, M., & Ronnander, C. (2009b). Integrated public use microdata series sample of the 1950 federal census. Accessed throgh usa.ipums.org/usa/.

U.S. Army Enlistment Records (1946). Army serial number electronic file, ca. 1938-1946. Electronic file from the National Archives and Records Administration, Washington, D.C.

U.S. Bureau of the Census (1880c). Tenth census of the United States, 1880, population schedule. Digital scans of original records in the National Archives, Washington, D.C., accessed through www.ancestry.com.

U.S. Bureau of the Census (1920b). Fourteenth census of the United States, 1920, population schedule. Digital scans of original records in the National Archives, Washington, D.C., accessed through www.ancestry.com.

U.S. Bureau of the Census (1930a). Fifteenth census of the United States, 1930, population schedule. Digital scans of original records in the National Archives, Washington, D.C., accessed through www.ancestry.com.

U.S. Public Health Service (1947). Vital statistics rates in the United States, 1900-1940. United States Government Printing Office.

U.S. War Department (1945). Department technical manual tm 12-305, machine records operation. United States War Department.

8 Tables and Figures

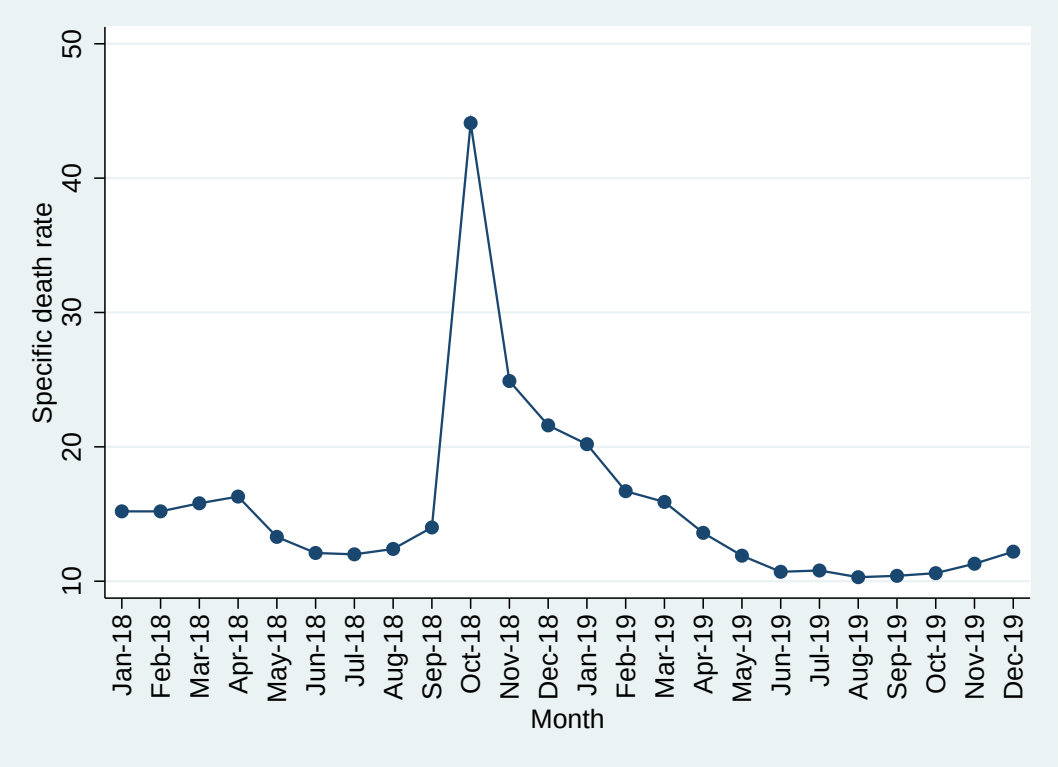

Figure 1: Deaths per 1,000 people by month, 1918-1919. Source: Vital Statistics Rates in the United States, 1900-1940 Table 2.

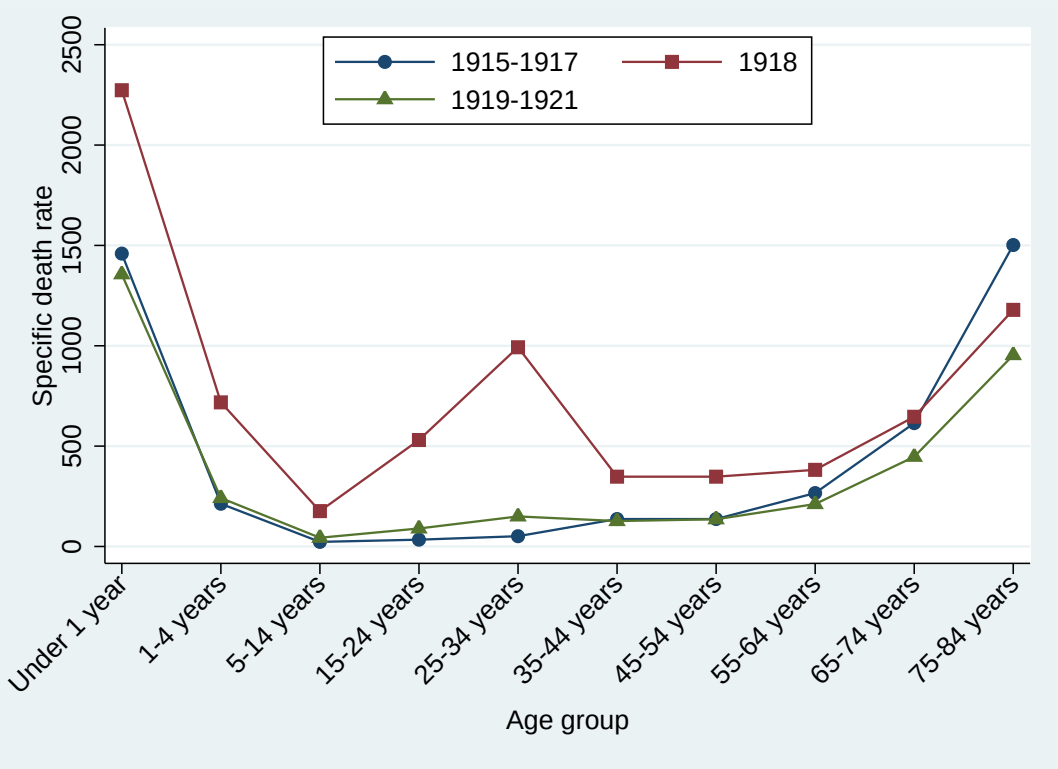

Figure 2: Deaths from influenza or pneumonia per 100,000 people by age group. Source: Vital Statistics Rates in the United States, 1900-1940 Table 14.

Table 1: Departure of 1919 male birth cohort outcomes from 1912-22 trend using modern census data and World War II enlistee data

	Results from Almond (2006)			Results using
Outcome	1960 Census	1970 Census	1980 Census	enlistee records
High school graduate	-0.021***	-0.020***	-0.014***	-0.028***
	(0.005)	(0.003)	(0.003)	(0.001)
Years of education	-0.150***	-0.176***	-0.117***	-0.132***
	(0.038)	(0.023)	(0.019)	(0.002)
Never attended high school	--	--	--	0.017***
	--	--	--	(0.001)
Height	--	--	--	-0.035***
	--	--	--	(0.007)
Weight	--	--	--	-0.371***
	--	--	--	(0.040)
Body Mass Index	--	--	--	-0.034***
	--	--	--	(0.006)
Observations	114,031	308,785	471,803	2,744,642

Robust standard errors are given in parentheses. * significant at 10%, ** significant at 5%, *** significant at 1%

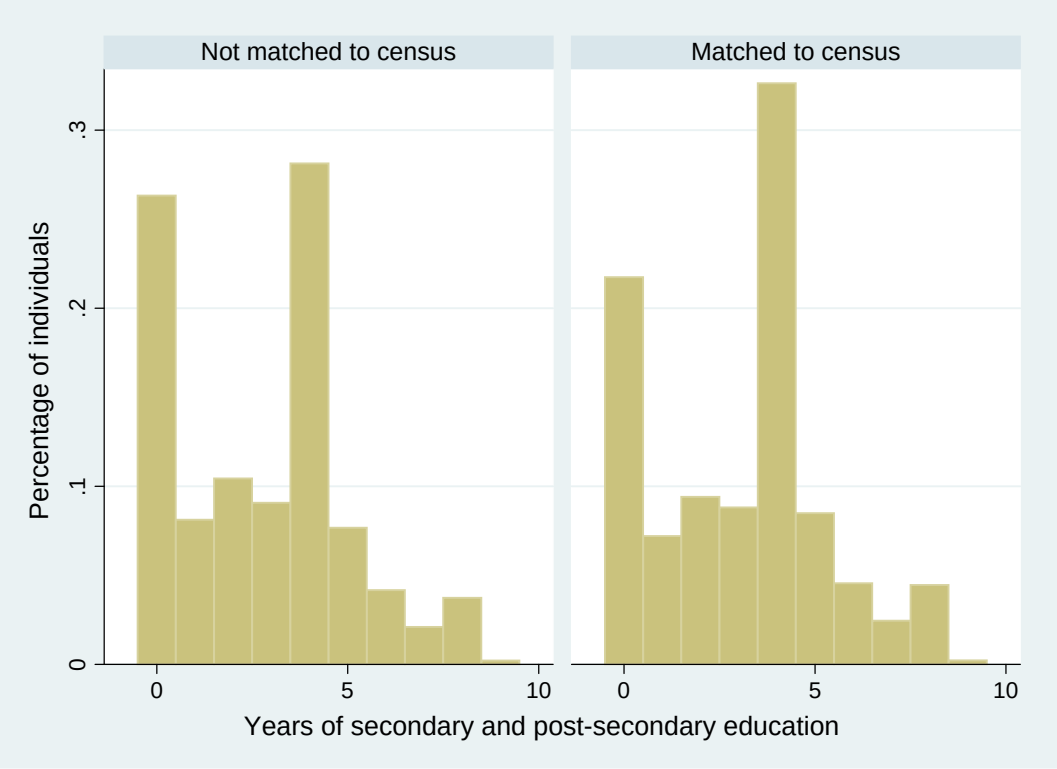

Figure 3: Distribution of educational attainment for enlistees by match outcome.

Table 2: Summary statistics for enlistees by match outcome

	Matched to federal census	Not matched to federal census
Number of individuals	13,173	14,139
Year of birth	1918.8	1918.7
	(3.2)	(3.2)
Year of enlistment	1941.8	1941.8
	(0.8)	(0.9)
Height	68.4	68.3
	(2.7)	(2.8)
Weight	150.2	149.9
	(21.3)	(21.7)
Body mass index	22.5	22.5
	(2.9)	(3.0)
Years of secondary and post-secondary schooling	3.0	2.8
	(2.2)	(2.3)
Percentage who are white	94.4%	90.1%
Percentage who migrated to a different state	13.5%	25.0%
Percentage who are sons of the household head	90.1%	--

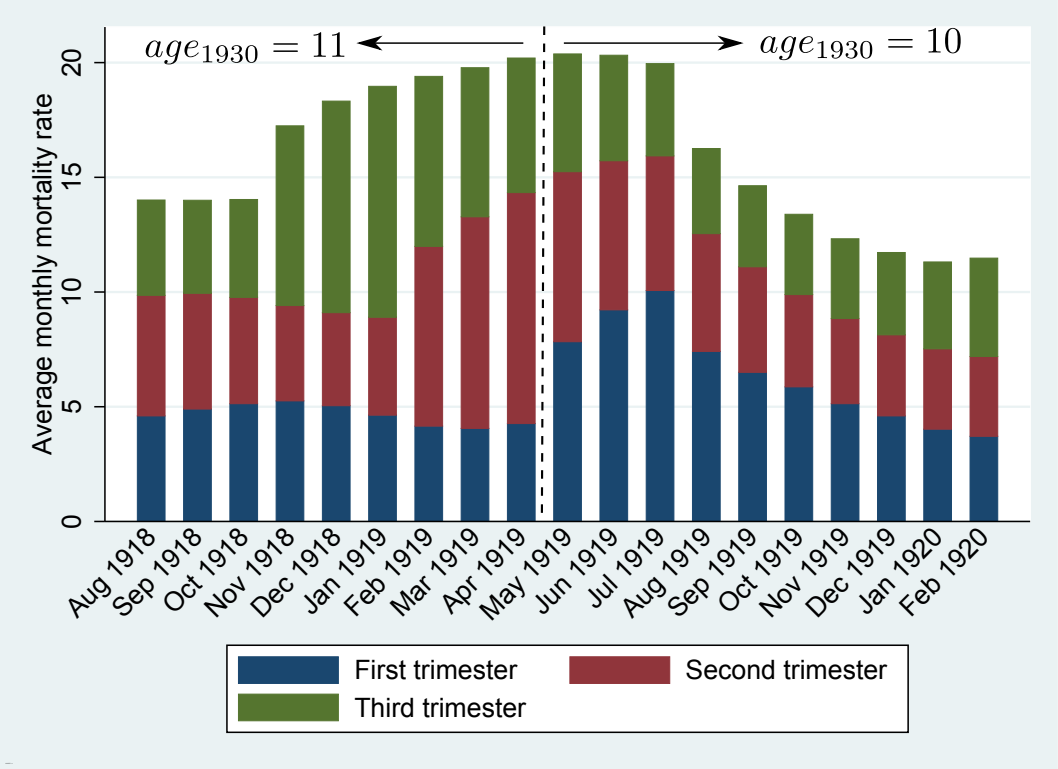

Figure 4: In utero exposure to the 1918 influenza pandemic by birthday and 1930 census age. Vertical axis measures average monthly mortality rate (deaths per 1,000 people) for the general population during the entire pregnancy, broken down by the relative contributions of mortality rates during the first, second and third trimesters. The monthly mortality rates used are the same as those given in Figure 1.

Table 3: Summary statistics enlistees in the regression sample

Variable	Mean	Standard deviation
Year of birth	1918.8	3.2
Height (inches)	68.4	2.7
Weight (pounds)	150.2	22.5
Body mass index	22.5	2.9
Years of secondary and post-secondary education	3.0	2.2
Number of people in household	6.4	2.3
Number of siblings	3.0	2.2
Number of brothers	1.6	1.5
Number of older siblings	1.5	1.6
Number of older brothers	0.8	1.1
Household head's income (1950 dollars)	2364.2	1136.5
Percentage living with father	94.5%	
Percentage living with mother	97.5%	
Percentage with an older sibling born in 1919	8.2%	
Percentage with a younger sibling born in 1919	6.9%	
Percentage with an older brother born in 1919	4.1%	
Percentage with a younger brother born in 1919	3.4%	

Table 4: Tobit estimates of the effects of sibling health on educational attainment, years of secondary and postsecondary schooling as dependent variable.

	Years of education			
Has older sibling born in 1919 (1=yes)	-0.016		-0.050	
	(0.111)		(0.112)	
Has younger sibling born in 1919 (1=yes)	0.300***		0.277**	
	(0.114)		(0.109)	
Number of older siblings	-0.174***		-0.172***	
	(0.021)		(0.019)	
Number of younger siblings	-0.363***		-0.357***	
	(0.020)		(0.018)	
Black (1=yes)	-1.223***	-1.223***	-1.103***	-1.103***
	(0.222)	(0.223)	(0.236)	(0.237)
Household head's log income	1.671***	1.672***	1.628***	1.629***
	(0.126)	(0.126)	(0.118)	(0.118)
Father present (1=yes)	-0.329**	-0.329**	-0.270	-0.269
	(0.157)	(0.158)	(0.167)	(0.167)
Mother present (1=yes)	0.676***	0.677***	0.681***	0.683***
	(0.165)	(0.166)	(0.157)	(0.158)
Has older brother born in 1919 (1=yes)		-0.089		-0.117
		(0.151)		(0.147)
Has older sister born in 1919 (1=yes)		0.059		0.017
		(0.139)		(0.140)
Has younger brother born in 1919 (1=yes)		0.263*		0.248*
		(0.147)		(0.138)
Has younger sister born in 1919 (1=yes)		0.335**		0.306*
		(0.165)		(0.162)
Number of older brothers		-0.171***		-0.171***
		(0.029)		(0.027)
Number of older sisters		-0.177***		-0.172***
		(0.036)		(0.031)
Number of younger brothers		-0.369***		-0.367***
		(0.033)		(0.029)
Number of younger sisters		-0.357***		-0.346***
		(0.027)		(0.022)
Born in 1919 (1=yes)			0.096	0.097
			(0.086)	(0.086)
N	8,647	8,647	9,759	9,759

* $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$

Robust standard errors clustered by birth state in parentheses, All regressions include a cubic in birth year and birth state fixed effects.

Table 5: Logit estimates of the effects of sibling health on educational attainment, graduated high school as dependent variable (1=graduated HS).

	High school grad			
Has older sibling born in 1919 (1=yes)	-0.113		-0.139	
	(0.104)		(0.105)	
Has younger sibling born in 1919 (1=yes)	0.249***		0.239***	
	(0.085)		(0.085)	
Number of older siblings	-0.134***		-0.126***	
	(0.021)		(0.017)	
Number of younger siblings	-0.268***		-0.268***	
	(0.019)		(0.018)	
Black (1=yes)	-1.094***	-1.094***	-0.927***	-0.928***
	(0.173)	(0.173)	(0.168)	(0.169)
Household head's log income	1.015***	1.016***	0.991***	0.992***
	(0.096)	(0.096)	(0.089)	(0.089)
Father present (1=yes)	-0.012	-0.012	0.055	0.056
	(0.176)	(0.176)	(0.174)	(0.174)
Mother present (1=yes)	0.543***	0.545***	0.586***	0.588***
	(0.119)	(0.121)	(0.126)	(0.127)
Has older brother born in 1919 (1=yes)		-0.205		-0.214
		(0.154)		(0.150)
Has older sister born in 1919 (1=yes)		-0.016		-0.063
		(0.116)		(0.118)
Has younger brother born in 1919 (1=yes)		0.163		0.164
		(0.100)		(0.104)
Has younger sister born in 1919 (1=yes)		0.330***		0.312***
		(0.114)		(0.110)
Number of older brothers		-0.129***		-0.123***
		(0.027)		(0.023)
Number of older sisters		-0.141***		-0.129***
		(0.031)		(0.026)
Number of younger brothers		-0.273***		-0.281***
		(0.034)		(0.032)
Number of younger sisters		-0.264***		-0.254***
		(0.024)		(0.018)
Born in 1919 (1=yes)			0.008	0.010
			(0.078)	(0.078)
N	8,647	8,647	9,759	9,759

* $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$

Robust standard errors clustered by birth state in parentheses, All regressions include a cubic in birth year and birth state fixed effects.

Table 6: Logit estimates of the effects of sibling health on educational attainment, attended high school as dependent variable (1=attended HS).

	Attended high school			
Has older sibling born in 1919 (1=yes)	0.001		-0.027	
	(0.096)		(0.098)	
Has younger sibling born in 1919 (1=yes)	0.392***		0.370***	
	(0.105)		(0.102)	
Number of older siblings	-0.126***		-0.125***	
	(0.018)		(0.017)	
Number of younger siblings	-0.245***		-0.239***	
	(0.016)		(0.015)	
Black (1=yes)	-0.758***	-0.758***	-0.715***	-0.716***
	(0.141)	(0.142)	(0.157)	(0.158)
Household head's log income	1.413***	1.418***	1.389***	1.392***
	(0.088)	(0.089)	(0.086)	(0.085)
Father present (1=yes)	-0.490***	-0.491***	-0.441***	-0.443***
	(0.141)	(0.139)	(0.164)	(0.163)
Mother present (1=yes)	0.397***	0.397***	0.399***	0.399***
	(0.145)	(0.145)	(0.138)	(0.138)
Has older brother born in 1919 (1=yes)		-0.105		-0.129
		(0.136)		(0.133)
Has older sister born in 1919 (1=yes)		0.117		0.080
		(0.126)		(0.129)
Has younger brother born in 1919 (1=yes)		0.375***		0.356***
		(0.126)		(0.112)
Has younger sister born in 1919 (1=yes)		0.407**		0.382**
		(0.167)		(0.164)
Number of older brothers		-0.105***		-0.112***
		(0.029)		(0.028)
Number of older sisters		-0.152***		-0.142***
		(0.033)		(0.031)
Number of younger brothers		-0.246***		-0.239***
		(0.025)		(0.022)
Number of younger sisters		-0.244***		-0.240***
		(0.025)		(0.021)
Born in 1919 (1=yes)			0.177**	0.178**
			(0.081)	(0.082)
N	8,639	8,639	9,759	9,759

* $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$

Robust standard errors clustered by birth state in parentheses, All regressions include a cubic in birth year and birth state fixed effects.

Table 7: Marginal effects ($\frac{dP(y=1)}{dx}$) of siblings on educational outcomes based on logit regressions.

	Graduated high	Attended high
Dependent variable:	school	school
Mean of dependent variable:	0.5236	0.7923
Has older sibling born in 1919	-0.0277	0.0002
	(0.0254)	(0.0165)
Has younger sibling born in 1919	0.0621	0.0612
	(0.0213)	(0.0149)
Number of older siblings	-0.0331	-0.0216
	(0.0052)	(0.0030)
Number of younger siblings	-0.0663	-0.0420
	(0.0047)	(0.0027)
Household head's log income	0.2508	0.2427
	(0.0236)	(0.0141)

Marginal effects based on the logit regressions in Table 4, column 1 and Table 5 column 1. Marginal effects are evaluated at the means of the independent variables. Standard errors given in parentheses.

Table 8: OLS estimates of the effects of sibling health on health, adult height in inches as dependent variable.

	Height (inches)			
Has older sibling born in 1919 (1=yes)	0.028		0.042	
	(0.110)		(0.107)	
Has younger sibling born in 1919 (1=yes)	0.069		0.057	
	(0.091)		(0.092)	
Number of older siblings	-0.073**		-0.075**	
	(0.028)		(0.030)	
Number of younger siblings	-0.131***		-0.121***	
	(0.026)		(0.024)	
Black (1=yes)	-0.527**	-0.533**	-0.585***	-0.589***
	(0.224)	(0.223)	(0.204)	(0.204)
Household head's log income	0.190*	0.190*	0.141	0.141
	(0.102)	(0.101)	(0.091)	(0.091)
Father present (1=yes)	0.233	0.232	0.258	0.259
	(0.227)	(0.228)	(0.205)	(0.205)
Mother present (1=yes)	0.491**	0.482**	0.524**	0.519**
	(0.228)	(0.231)	(0.214)	(0.217)
Has older brother born in 1919 (1=yes)		0.042		0.025
		(0.155)		(0.156)
Has older sister born in 1919 (1=yes)		0.009		0.056
		(0.147)		(0.130)
Has younger brother born in 1919 (1=yes)		0.156		0.161
		(0.150)		(0.143)
Has younger sister born in 1919 (1=yes)		-0.006		-0.041
		(0.155)		(0.154)
Number of older brothers		-0.091**		-0.090**
		(0.039)		(0.042)
Number of older sisters		-0.051		-0.057
		(0.037)		(0.035)
Number of younger brothers		-0.078*		-0.078**
		(0.043)		(0.038)
Number of younger sisters		-0.187***		-0.166***
		(0.030)		(0.028)
Born in 1919 (1=yes)			-0.146*	-0.150*
			(0.075)	(0.076)
R^2	0.05	0.05	0.05	0.05
N	8,647	8,647	9,759	9,759

* $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$

Robust standard errors clustered by birth state in parentheses, All regressions include a cubic in birth year and birth state fixed effects.

Table 9: OLS estimates of the effects of sibling health on health, body mass index as dependent variable.

	Body mass index			
Has older sibling born in 1919 (1=yes)	0.080		0.073	
	(0.093)		(0.100)	
Has younger sibling born in 1919 (1=yes)	-0.040		-0.049	
	(0.090)		(0.097)	
Number of older siblings	-0.022		-0.017	
	(0.016)		(0.016)	
Number of younger siblings	-0.026		-0.021	
	(0.018)		(0.017)	
Black (1=yes)	0.420***	0.425***	0.475***	0.479***
	(0.133)	(0.135)	(0.158)	(0.160)
Household head's log income	0.006	0.007	-0.016	-0.014
	(0.114)	(0.113)	(0.103)	(0.102)
Father present (1=yes)	0.179	0.179	0.183	0.181
	(0.212)	(0.210)	(0.190)	(0.188)
Mother present (1=yes)	0.231	0.233	0.157	0.159
	(0.145)	(0.144)	(0.154)	(0.154)
Has older brother born in 1919 (1=yes)		0.131		0.132
		(0.118)		(0.119)
Has older sister born in 1919 (1=yes)		0.041		0.023
		(0.139)		(0.141)
Has younger brother born in 1919 (1=yes)		-0.057		-0.064
		(0.126)		(0.133)
Has younger sister born in 1919 (1=yes)		-0.028		-0.040
		(0.117)		(0.120)
Number of older brothers		0.007		0.010
		(0.026)		(0.025)
Number of older sisters		-0.059**		-0.050**
		(0.027)		(0.024)
Number of younger brothers		-0.048*		-0.046*
		(0.026)		(0.025)
Number of younger sisters		-0.003		0.006
		(0.027)		(0.025)
Born in 1919 (1=yes)			-0.067	-0.063
			(0.100)	(0.101)
R^2	0.05	0.05	0.04	0.05
N	8,528	8,528	9,629	9,629

* $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$

Robust standard errors clustered by birth state in parentheses, All regressions include a cubic in birth year and birth state fixed effects.

Table 10: Coefficients from falsification tests using placebo influenza cohorts.

	Influenza cohort defined as being born in:		
	1917	1919	1921
Panel A: Tobit coefficients, years of secondary and postsecondary education as dependent variable			
Older sibling born in flu cohort	0.012	-0.016	-0.130
(1=yes)	(0.105)	(0.111)	(0.172)
Younger sibling born in flu	-0.051	0.300***	-0.022
cohort (1=yes)	(0.161)	(0.114)	(0.088)
Panel B: Logit coefficients, graduated high school as dependent variable			
Older sibling born in flu cohort	-0.021	-0.113	-0.159
(1=yes)	(0.083)	(0.104)	(0.114)
Younger sibling born in flu	-0.100	0.249***	-0.028
cohort (1=yes)	(0.098)	(0.085)	(0.063)
Panel C: Logit coefficients, graduated high school as dependent variable			
Older sibling born in flu cohort	0.016	0.001	-0.218
(1=yes)	(0.096)	(0.096)	(0.169)
Younger sibling born in flu	-0.019	0.392***	-0.035
cohort (1=yes)	(0.122)	(0.105)	(0.079)

Robust standard errors clustered by birth state in parentheses, * significant at 10%; ** significant at 5%; *** significant at 1%. Regressions follow the same specifications as columns 1 of Table 3, Table 4 and Table 5 for Panels A, B and C, respectively.

A Matching Enlistment Records to the Federal Census

The process of matching individuals from the enlistment records to the federal census was done by hand for an initial sample roughly 20,000 observations. This initial sample used to create an algorithm to automate the matching for the remainder of the sample. This appendex provides a discussion of the criteria by which matches can be evaluated and details of the matching algorithm.

Individuals can be searched for in an online index of the 1930 federal census on the basis of name, year of birth, state of residence, state of birth, gender and race. In practice, gender and race prove less useful for matching than one would initially expect. Unlike the other characteristics, these two variables are each recorded in the census using a single character: gender is recorded as either "M" or "F" for male or female, race is typically recorded as either "W" or "N" for white or negro. The difficulty this presents is that having only a single character makes poor handwriting far more problematic. Unlike name or state, there is no context to help decipher handwriting. Consequently, incorrect information for race or gender in the electronic index is not uncommon. If you include race and gender in the search criteria, an otherwise exact match with a incorrectly transcribed race or gender will appear after individuals with noticeably worse matches for name or birth year who have the correct race or gender. For this reason, race and gender are not included as search criteria (although they can still be used to choose between multiple potential matches).

This leaves name, birth year, birth state and state of residence as the search criteria. The enlistment records provide birth state and state of residence at the time of enlistment. Assuming that the state of residence at the time of enlistment is the same as the state of residence in 1930 is rather restrictive. Consequently, only state of birth is used in the search of the federal census. Thus the final set of search criteria are name, birth year and birth state. When these search criteria are entered in ancestry.com, the site returns a list of potential matches ordered from the best match to the worst match. The matching criteria are not strictly enforced. Some results may have slightly different names, some may have slightly different birth years and some may have a different birth state. The purpose of the matching algorithm is to determine whether

there is a sufficiently close match among these search results and, if so, whether that match is unique.

The first step in finding the match results is to use a web script to scrape the information for the top five census results. The information about each result that can be scraped includes the full name of the individual in the census, the first names of the individual's parents, the city, county and state of residence in 1930, birth year, birth state, relation to the head of the household, and the url for the webpage containing the complete transcribed information and a link image of the census manuscript page for the individual. Given these variables, each match can be assessed on how closely the first name matches the first name in the enlistment records, how closely the last names match, how close the birth year is, whether the birth state matches the birth state in the enlistment records, and whether the state of residence in 1930 matches the birth state in the enlistment records. Assessing whether the states match or whether the birth years match is straightforward. Whether the states match can be captured by a simple binary variable indicating yes or no. How well the birth year matches can measured simply by the difference in birth years between the enlistment records and the federal census.[14] Assessing how closely names match is more difficult.

There are two approaches taken for measuring how well the names match. The first is a very coarse measure based on the Phonex algorithm designed to determine whether two words are phonetically equivalent. The Phonex algorithm incorporates principles from both the Soundex and Metaphone methods of name matching. An overview of the algorithm and its performance relative to other approached to coding names phonetically is provided by Lait & Randell (1996). The Phonex algorithm is used to convert the first name and the last name into their respective Phonex codes and compare these codes to the codes for each of the search results to determine whether the names are phonetically equivalent.

This approach provides an easy way to eliminate matches that are clearly incorrect names. However, it cannot distinguish between names that have similar structures but are clearly different. Consider the examples provided in Table 11 of pairs of last names that produce the

[14]It should be noted that while ancestry.com returns the birth year in the search results, the actual information provided in the census is the individual's age at the time of the census (April 1st, 1930). What ancestry.com is reporting as the birth year is simply 1930 minus this age. This imputed birth year may be off by one year depending whether the individual's birthday falls before or after April 1st.

Table 11: Last names coded using the Phonex algorithm.

Name A	Name B	Phonex Code A	Phonex Code B	Damerau-Levenshtein Distance between A and B
Harland	Olin	A450	A450	5
Armando	Herman	A500	A500	4
Francis	Franklin	F652	F652	4
Garland	Glenn	G450	G450	5
Michael	Mitchell	M240	M240	3
Thomas	Thurman	T500	T500	5
Fielding	Feilding	F350	F350	1
Winfield	Winford	W513	W513	4

same Phonex code. While the Phonex algorithm provides an efficient way to quickly eliminate drastically different names, it will not help identify slight mispellings or otherwise small differences between names. This leads to the second measure for whether names match, the DamerauLevenshtein distance.

The Damerau-Levenshtein distance is a measure of the distance between two strings based on the number of operations it requires to transform one string into the other. The allowed operations are the addition of a character, the substraction of a character, the replacement of a character with another character, and the transposition of two characters. The Damerau-Levenshtein distance in the minimum number of these four operations required to get from one string to the other. These distances are also shown in Table 11. While the Damerau-Levenshtein difference itself is a useful measure of how close two strings are to one another, it does not take into account the overall length of the string. Two four-letter names that differ by two characters are assigned the same distance from one another as two ten-letter names that differ by two characters. To take into account the distance between two names relative to their overall length, I also construct a normalized version of the Damerau-Levenshtein difference equal to the Damerau-Levenshtein distance divided by the average lenght of the two names.

The final set of variables that can be used to identify matches are summarized in **??**. To

evaluate matches, I construct a numerical score of the form

$$\theta(X_{i,j}) = \beta_1 X^1_{i,j} + \beta_2 X^2_{i,j} + ... + \beta_k X^k_{i,j} \tag{3}$$

where $\theta(X_{i,j})$ is the match score for census search result j for enlistee i based on the k different matching variables. To estimate the appropriate coefficients to use when constructing this score, I use the set of observations matched hand for which there was a unique match found. Letting $y_{i,j}$ be a indicator variable equal to one if census search result j is determined to be a match for individual i and zero if it is not a match, every enlistee in this set of hand matched observations for which a unique match is found satisfies the following property:

$$\sum_{j=1}^{5} y_{i,j} = 1 \tag{4}$$

This property suggests a very natural way to estimate the parameters for $\theta(X_{i,j})$. Essentially, this is a situation where there is a multinomial discrete choice model in which the regressors vary across alternatives (the five different census search results) but the coefficients do not. By restricting the sample to those individuals with one and only one unique match, this can be modeled as a conditional logit model:

$$y_{i,j} = \frac{e^{\theta(X_{i,j})}}{\sum_{l=1}^{5} e^{\theta(X_{i,l})}} \qquad \text{s.t.} \ \sum_{j=1}^{5} y_{i,j} = 1 \tag{5}$$

The estimated coefficients from this model provide the function

$$\hat{\theta}(X_{i,j}) = \hat{\beta}_1 X^1_{i,j} + ... + \hat{\beta}_k X^k_{i,j} \tag{6}$$

that can be used to calculate match scores for all of the observations in the sample. The conditional logit regression results based used to construct $\hat{\theta}(X_{i,j})$ are give in Table 13.

The final step is to use the cases of multiple matches and no matches to establish cutoffs for the match score above which a census search result can be considered a match. Let $z_i(X_{i,1}, ..., X_{i,5})$

be an indicator variable for the matching outcome of enlistee i:

$$z_i(X_{i,1}, ..., X_{i,5}) = \begin{cases} 0 & \text{if there are zero matches} \\ 1 & \text{if there is a unique match} \\ 2 & \text{if there are multiple matches} \end{cases} \tag{7}$$

Letting $\overline{\theta}$ represent the score cutoff above which a census search result is considered a match and ordering the five census search results from highest score to lowest score, z_i can be expressed as:

$$z_i(X_{i,1}, ..., X_{i,5}) = \begin{cases} 0 & \text{if } \hat{\theta}(X_{i,1}) < \overline{\theta} \\ 1 & \text{if } \hat{\theta}(X_{i,1}) \geq \overline{\theta} \text{ and } \hat{\theta}(X_{i,2}) < \overline{\theta} \\ 2 & \text{if } \hat{\theta}(X_{i,1}) \geq \overline{\theta} \text{ and } \hat{\theta}(X_{i,2}) \geq \overline{\theta} \end{cases} \tag{8}$$

Since z_i is observed in the hand-matched data, the full set of hand-matched data can be used to estimate the value of $\overline{\theta}$ based on preferences over the types of matching errors. Choosing a higher value for $\overline{\theta}$ will reduce the probability of declaring census search results matches when in fact they are not and the probability of declaring multiple matches when there is actually a unique match. However, the penalty for choosing a higher value of $\overline{\theta}$ to reduce false matches is a lower probability of identifying a unique match when one exists.

Figure 5 and Figure 6 show the tradeoffs associated with different levels for $\overline{\theta}$. At low levels of the cutoff, increasing the cutoff increases the number of correctly identified unique matches in the hand-matched data rises but so does the number of false unique matches (observations that actually have either multiple matches or no matches). However, for cutoff values above -1, the number of correctly identified unique matches continues to rise with $\overline{\theta}$ while the number of false matches begins to decline. The ratio of correct unique matches to false unique matches continues to rise up to a cutoff value of 3.9 where the ratio reaches 4.6.

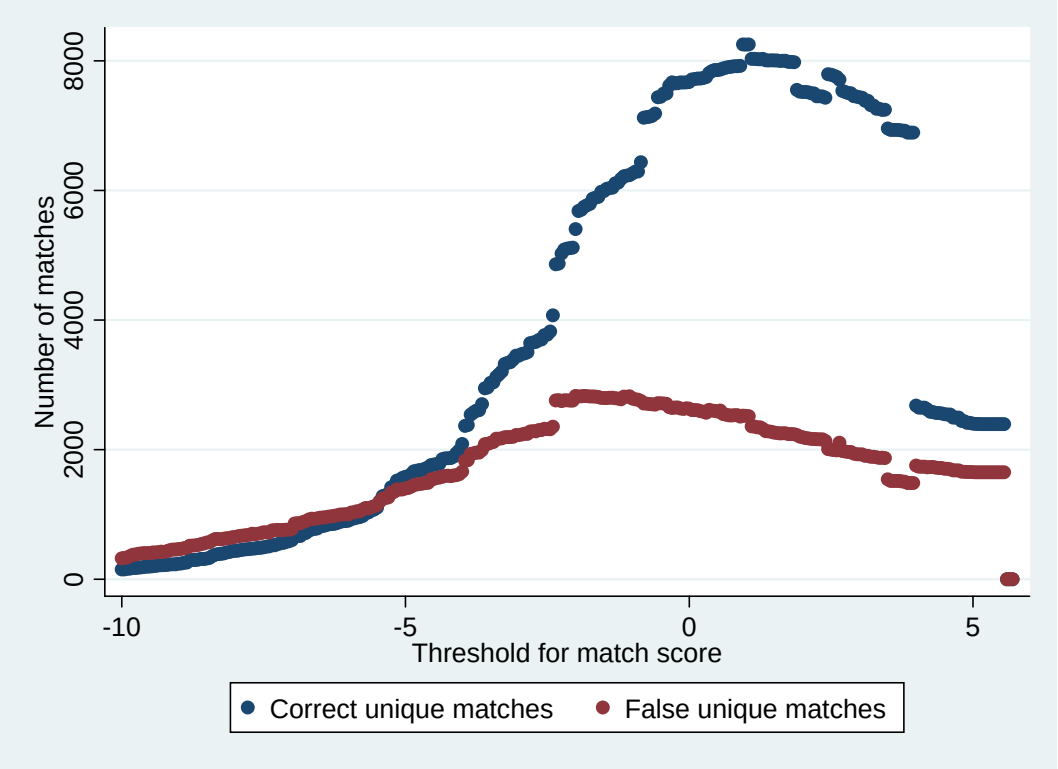

Figure 5: Correct and false unique matches by threshold for match score.

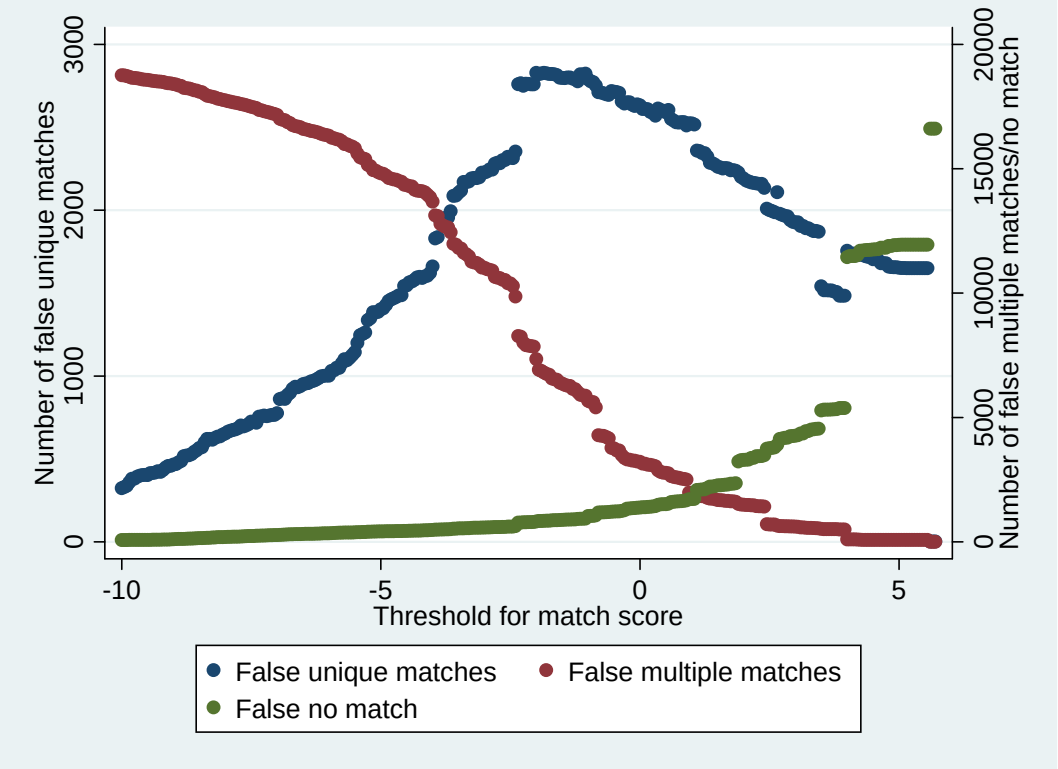

Figure 6: False matches by type by threshold for match score.

Table 12: Variables for evaluating matches between enlistment and census records.

Variable	Definition	Type
Birth state	Equal to 1 if birth state in census matches birth state in enlistment record, 0 otherwise	Binary
Residence state A	Equal to 1 if state of residence in census matches state of residence in enlistment record, 0 otherwise	Binary
Residence state B	Equal to 1 if state of residence in census matches birth state in enlistment record, 0 otherwise	Binary
Age difference	Absolute value of difference between enlistment record birth year and imputed census birth year (1930 - census age)	Discrete
Age difference squared	Square of the variable above	Discrete
Phonex first name	Equal to 1 if census first name and enlistment record first name have identical phonex codes, 0 otherwise	Binary
Phonex last name	Equal to 1 if census last name and enlistment record last name have identical phonex codes, 0 otherwise	Binary
Damerau-Levenshtein first name	Number of operations required to transform census first name into enlistment record first name	Discrete
Damerau-Levenshtein last name	Number of operations required to transform census last name into enlistment record last name	Discrete
Normalized D-L first name	Damerau-Levenshtein first name distance divided by the average length of the first names	Continuous
Normalized D-L last name	Damerau-Levenshtein last name distance divided by the average length of the last names	Continuous

Table 13: Conditional logit results to estimate parameters for the match score using the set of hand-matched enlistees with a unique census match.

	Coefficient (standard error)	Variable mean in regression sample (standard deviation)
Absolute value of difference in age	-1.6127***	2.0462
	(0.0425)	(2.7338)
Difference in age squared	0.0185***	11.6603
	(0.0005)	(126.2093)
First name phonex match (1=yes)	-0.6652*	0.5500
	(0.3454)	(0.4975)
Last name phonex match (1=yes)	-0.1350	0.9063
	(0.3541)	(0.2914)
Normalized first name Damerau-Levenshtein distance	-9.9404***	0.3671
	(0.5368)	(0.4089)
Normalized first name Damerau-Levenshtein distance	4.3198***	0.0079
x First name phonex match	(0.7393)	(0.0560)
Normalized last name Damerau-Levenshtein distance	-6.6829***	0.0798
	(0.7415)	(0.2042)
Normalized last name Damerau-Levenshtein distance	-5.7693***	0.0307
x Last name phonex match	(0.9339)	(0.2042)
Census birth state matches enlistment	1.3759***	0.7075
birth state (1=yes)	(0.1288)	(0.4549)
Census residence state matches enlistment	2.9010***	0.6937
birth state (1=yes)	(0.1429)	(0.4610)
Census residence state matches enlistment	2.1052***	0.6245
residence state (1=yes)	(0.1293)	(0.4842)
Observations	51814	
Pseudo R-squared	0.8971	

* significant at 10%; ** significant at 5%; *** significant at 1%

B Additional Tables and Figures

Table 14: Tobit estimates of the effects of sibling health on educational attainment with birth year fixed effects, years of secondary and postsecondary schooling as dependent variable.

	Years of education	
Has older sibling born in 1919 (1=yes)	-0.033	
	(0.109)	
Has younger sibling born in 1919 (1=yes)	0.264**	
	(0.116)	
Number of older siblings	-0.172***	
	(0.021)	
Number of younger siblings	-0.361***	
	(0.020)	
Black (1=yes)	-1.214***	-1.095***
	(0.224)	(0.238)
Household head's log income	1.668***	1.626***
	(0.126)	(0.118)
Father present (1=yes)	-0.339**	-0.278*
	(0.159)	(0.167)
Mother present (1=yes)	0.665***	0.673***
	(0.165)	(0.159)
Has older brother born in 1919 (1=yes)		-0.135
		(0.143)
Has older sister born in 1919 (1=yes)		0.000
		(0.140)
Has younger brother born in 1919 (1=yes)		0.211
		(0.136)
Has younger sister born in 1919 (1=yes)		0.272
		(0.167)
Number of older brothers		-0.169***
		(0.027)
Number of older sisters		-0.171***
		(0.031)
Number of younger brothers		-0.365***
		(0.030)
Number of younger sisters		-0.346***
		(0.022)
N	8,647	9,759

* $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$

Robust standard errors clustered by birth state in parentheses, All regressions include birth state and birth year fixed effects.

Table 15: Logit estimates of the effects of sibling health on educational attainment with birth year fixed effects.

	High school grad		Attended high school	
Has older sibling born in 1919 (1=yes)	-0.146		-0.014	
	(0.100)		(0.096)	
Has younger sibling born in 1919 (1=yes)	0.216**		0.354***	
	(0.087)		(0.109)	
Number of older siblings	-0.132***		-0.125***	
	(0.021)		(0.018)	
Number of younger siblings	-0.267***		-0.243***	
	(0.019)		(0.016)	
Black (1=yes)	-1.091***	-0.924***	-0.756***	-0.714***
	(0.176)	(0.171)	(0.143)	(0.159)
Household head's log income	1.020***	0.996***	1.414***	1.393***
	(0.097)	(0.090)	(0.089)	(0.086)
Father present (1=yes)	-0.023	0.046	-0.491***	-0.444***
	(0.178)	(0.176)	(0.142)	(0.165)
Mother present (1=yes)	0.537***	0.582***	0.395***	0.397***
	(0.122)	(0.129)	(0.147)	(0.140)
Has older brother born in 1919 (1=yes)		-0.251*		-0.146
		(0.142)		(0.131)
Has older sister born in 1919 (1=yes)		-0.091		0.069
		(0.119)		(0.129)
Has younger brother born in 1919 (1=yes)		0.129		0.318***
		(0.105)		(0.112)
Has younger sister born in 1919 (1=yes)		0.283**		0.346**
		(0.112)		(0.170)
Number of older brothers		-0.121***		-0.111***
		(0.024)		(0.027)
Number of older sisters		-0.128***		-0.142***
		(0.026)		(0.031)
Number of younger brothers		-0.278***		-0.237***
		(0.032)		(0.022)
Number of younger sisters		-0.254***		-0.240***
		(0.018)		(0.021)
N	8,636	9,748	8,639	9,759

* $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$

Robust standard errors clustered by birth state in parentheses, All regressions include birth state and birth year fixed effects.

Table 16: OLS estimates of the effects of sibling health on own health with birth year fixed effects.

	Height (inches)		Body mass index		Weight (pounds)	Weight (pounds)
Has older sibling born in 1919 (1=yes)	0.009		0.073		0.376	
	(0.110)		(0.096)		(0.561)	
Has younger sibling born in 1919 (1=yes)	0.051		-0.060		-0.160	
	(0.092)		(0.091)		(0.618)	
Number of older siblings	-0.071**		-0.023		-0.483***	
	(0.028)		(0.016)		(0.137)	
Number of younger siblings	-0.131***		-0.025		-0.809***	
	(0.026)		(0.019)		(0.107)	
Black (1=yes)	-0.518**	-0.581***	0.423***	0.482***	0.863	0.929
	(0.225)	(0.205)	(0.135)	(0.161)	(1.031)	(1.105)
Household head's log income	0.188*	0.140	0.006	-0.014	0.867	0.540
	(0.102)	(0.091)	(0.113)	(0.102)	(0.836)	(0.736)
Father present (1=yes)	0.214	0.243	0.185	0.186	2.081	2.237
	(0.225)	(0.203)	(0.213)	(0.189)	(1.785)	(1.441)
Mother present (1=yes)	0.492**	0.520**	0.225	0.154	3.802**	3.215**
	(0.230)	(0.219)	(0.141)	(0.151)	(1.608)	(1.571)
Has older brother born in 1919 (1=yes)		0.005		0.119		0.627
		(0.156)		(0.122)		(0.859)
Has older sister born in 1919 (1=yes)		0.042		0.023		0.235
		(0.130)		(0.142)		(0.949)
Has younger brother born in 1919 (1=yes)		0.146		-0.082		-0.004
		(0.142)		(0.127)		(0.785)
Has younger sister born in 1919 (1=yes)		-0.062		-0.064		-0.571
		(0.153)		(0.127)		(0.959)
Number of older brothers		-0.086**		0.010		-0.380**
		(0.042)		(0.025)		(0.188)
Number of older sisters		-0.058		-0.051**		-0.560**
		(0.035)		(0.024)		(0.215)
Number of younger brothers		-0.078**		-0.045*		-0.717***
		(0.038)		(0.025)		(0.205)
Number of younger sisters		-0.166***		0.006		-0.732***
		(0.028)		(0.026)		(0.198)
R^2	0.05	0.05	0.05	0.05	0.03	0.03
N	8,647	9,759	8,528	9,629	8,528	9,629

* $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$

Robust standard errors clustered by birth state in parentheses, All regressions include birth state and birth year fixed effects.

Table 17: Coefficients from falsification tests for the tobit results using placebo influenza cohorts, years of secondary and postsecondary education as the dependent variable.

	Influenza cohort defined as being born in:					
	1919	1917	1921	1919	1917	1921
	(1)	(2)	(3)	(4)	(5)	(6)
Includes birth year fixed effects	no	no	no	yes	yes	yes
Has older sibling born in flu cohort (1=yes)	-0.016	0.012	-0.130	-0.033	0.006	-0.192
	(0.111)	(0.105)	(0.172)	(0.109)	(0.104)	(0.183)
Has younger sibling born in flu cohort (1=yes)	0.300***	-0.051	-0.022	0.264***	-0.058	-0.021
	(0.114)	(0.161)	(0.088)	(0.116)	(0.162)	(0.088)
Number of older siblings	-0.174***	-0.175***	-0.173***	-0.172***	-0.174***	-0.171***
	(0.021)	(0.017)	(0.020)	(0.021)	(0.017)	(0.020)
Number of younger siblings	-0.363***	-0.350***	-0.349***	-0.361***	-0.349***	-0.348***
	(0.020)	(0.021)	(0.019)	(0.020)	(0.021)	(0.019)
Household head's log income	1.671***	1.616***	1.628***	1.668***	1.613***	1.625***
	(0.126)	(0.122)	(0.118)	(0.126)	(0.122)	(0.271)

Robust standard errors clustered by birth state in parentheses, * significant at 10%; ** significant at 5%; *** significant at 1%. All regressions follow use the same specification as column 1 of Table 3, controlling for race, presence of the father and mother and birth state fixed effects.

Table 18: Coefficients from falsification tests for the logit results using placebo influenza cohorts, graduated high school as dependent variable.

	Influenza cohort defined as being born in:					
	1919	1917	1921	1919	1917	1921
	(1)	(2)	(3)	(4)	(5)	(6)
Includes birth year fixed effects	no	no	no	yes	yes	yes
Has older sibling born in flu cohort (1=yes)	-0.113	-0.021	-0.159	-0.146	-0.014	-0.277**
	(0.104)	(0.083)	(0.114)	(0.100)	(0.079)	(0.122)
Has younger sibling born in flu cohort (1=yes)	0.249***	-0.100	-0.028	0.216**	-0.124	-0.012
	(0.085)	(0.098)	(0.063)	(0.087)	(0.101)	(0.065)
Number of older siblings	-0.134***	-0.132***	-0.120***	-0.132***	-0.132***	-0.117***
	(0.021)	(0.016)	(0.017)	(0.021)	(0.015)	(0.017)
Number of younger siblings	-0.268***	-0.262***	-0.256***	-0.267***	-0.261***	-0.257***
	(0.019)	(0.018)	(0.017)	(0.019)	(0.018)	(0.018)
Household head's log income	1.015***	0.979***	0.979***	1.020***	0.984	0.985***
	(0.096)	(0.094)	(0.088)	(0.097)	(0.094)	(0.089)

Robust standard errors clustered by birth state in parentheses, * significant at 10%; ** significant at 5%; *** significant at 1%. All regressions follow use the same specification as column 1 of Table 4, controlling for race, presence of the father and mother and birth state fixed effects.

Table 19: Coefficients from falsification tests for the logit results using placebo influenza cohorts, attended high school as dependent variable.

	Influenza cohort defined as being born in:					
	1919	1917	1921	1919	1917	1921
	(1)	(2)	(3)	(4)	(5)	(6)
Includes birth year fixed effects	no	no	no	yes	yes	yes
Has older sibling born in flu cohort (1=yes)	0.001	0.016	-0.218	-0.014	-0.002	-0.263
	(0.096)	(0.096)	(0.169)	(0.096)	(0.096)	(0.182)
Has younger sibling born in flu cohort (1=yes)	0.392***	-0.019	-0.035	0.354***	-0.023	-0.046
	(0.105)	(0.122)	(0.079)	(0.109)	(0.117)	(0.079)
Number of older siblings	-0.126***	-0.128***	-0.121***	-0.125***	-0.128***	-0.121***
	(0.018)	(0.016)	(0.019)	(0.018)	(0.016)	(0.019)
Number of younger siblings	-0.245***	-0.228***	-0.220***	-0.243***	-0.227***	-0.219***
	(0.016)	(0.016)	(0.014)	(0.016)	(0.017)	(0.014)
Household head's log income	1.413***	1.388***	1.361***	1.414***	1.391***	1.365***
	(0.088)	(0.089)	(0.095)	(0.089)	(0.089)	(0.096)

Robust standard errors clustered by birth state in parentheses, * significant at 10%; ** significant at 5%; *** significant at 1%. All regressions follow use the same specification as column 1 of Table 5, controlling for race, presence of the father and mother and birth state fixed effects.