

## *Online Appendix*

### *Childhood Health and Human Capital: New Evidence from Genetic Brothers in Arms*

#### DATA SOURCES

##### *World War II Enlistment Records and the Matching of Brothers*

The World War II enlistment records were obtained from the National Archives and Records Administration (NARA) as an electronic file. These electronic records were converted from the Army Serial Number microfilm of computer punchcards by the NARA and include records for roughly nine million men and women who enlisted in the United States Army between 1938 and 1946. The records are not complete both due to missing records for certain ranges of serial numbers and because several thousand records could not be interpreted by the NARA's scanning system.

The relevant variables reported in the enlistment records include: serial number, name, state and county of residence, place of enlistment, date of enlistment, military grade, military branch, nativity, year of birth, race, education, civilian occupation, marital status, height, and weight. Not all variables were reported in all years. The most important change in the reporting over time for the purposes of this article was the exclusion of height and weight information after 1943. Consequently, this study is restricted to individuals enlisting between 1938 and 1943. In some records, what replaced the height and weight information was actually the enlistee's score on the Army General Classification Test (AGCT), a test of cognitive ability. While this article focuses on the relationship between height and educational attainment, a similar study comparing differences in educational attainment to differences in AGCT scores between brothers would certainly be worthwhile.

A further complication with the height and weight variables is that the reporting of those variables was inconsistent. The NARA notes that instructions for the use of the height and weight fields changed during the war and that some cards contain information on military occupation in the height and weight fields. However, there is no way to know for certain which cards report height and weight and which cards use the fields to report something else. In an attempt to restrict the sample to records reporting height and weight, I discard observations for which the stated height and weight imply an unrealistic body mass index for an individual. I compute the body mass index based on the stated height and weight and discard observations with a body mass index (BMI) of less than 15 (below which individuals are considered to be exhibiting starvation) or greater than 60 (above which individuals are considered hyper-obese). Despite these precautions, there may still be observations in the sample for which the height and weight fields do not actually contain information on height and weight.

For the purposes of documenting the secular trends in height and educational attainment, the sample is further restricted to include only those enlistees who were assigned the rank of private. Many of the individuals assigned higher ranks have ages that correspond to having served in World War I and are re-enlisting as officers for World War II (explaining their higher ranks). The army would discard an individual's old enlistment card and create a new card upon re-enlistment. Consequently, these officers from World War I re-enlisting to serve in World War II have an enlistment record that appears just like that of a draftee with the exception of the rank. These officers create a sample selection problem when it comes to documenting the secular

trends in both height and education. Officers are on average significantly taller and more educated relative to other members of the army. The birth cohort that corresponds to World War I veterans has a disproportionate number of officers in the enlistment sample and therefore appears significantly taller and more educated than either the cohort before them or after them. To keep samples of the birth cohorts comparable across birthyears, I restrict the sample to privates.

The following procedure was used to create the sample of brothers from the full enlistment records. First, as previously described, all individuals with suspect height and weight data were discarded. Next, the individuals were sorted by last name, state of birth, state and county of residence, and then age. Potential brothers were identified as individuals sharing the same last name, state of birth, and state and county of residence, and within three years of each other in age. Every tenth set of potential brothers was kept, creating a 10 percent sample of the enlistment records.

A Perl script was then used to search for every potential brother in the 10 percent sample in either the 1930 or the 1920 federal census. If all individuals in a group of potential brothers were born prior to 1920, the 1920 federal census was used for all brothers in the group. If any of the individuals in a group of potential brothers was born in 1920 or later, the 1930 federal census was used for all brothers in the group. For each individual, the Perl script searches ancestry.com's online database of census records using the individual's first and last names, state of birth, and birthyear and returns the location of the person considered to be the best match in the federal census. All individuals in a group of potential brothers are then sorted by county of residence in the federal census and parents' names. Only potential brothers living in the same county in the federal census with identical parents' names are kept.

Next, these remaining individuals are then searched for by hand in the ancestry.com database to confirm that they have a unique match in the database. If there is not a unique match (multiple individuals have the same name and were born within one year of the enlistee's birthyear or no individuals exactly match both the name and birthyear of the enlistee record), the individual is discarded. For the remaining potential brothers with unique matches, images of the original census manuscripts containing the individuals are downloaded. From these images, it is possible to determine whether the potential brothers truly lived in the same household. If so, they are recorded as a confirmed match and information on the father and household structure are transcribed from the census image. If not, the individuals are dropped from the dataset. Roughly one-third of the potential brothers have a unique match in the federal census. Of these uniquely matched potential brothers, roughly one quarter are actually in the same household as one of the other uniquely matched potential brothers.

The nature of the enlistment data and the process of linking brothers raises several different sample selection problems. For the enlistment records themselves, there are concerns that the enlistees may not be representative of the general population. Potential enlistees could be rejected both for health issues and for inadequate education, the two main dimensions of interest in this study. The linking procedure leads to a sample of men underrepresentative of geographically mobile individuals and of smaller families. The next sections will offer additional statistics to evaluate the extent of these selection problems.

### *Selection Issues Related to the Army Enlistment Standards*

While using military enlistment records from World War II has the advantage of including a large number of conscripted individuals randomly selected from the population of adult males, the criteria for being eligible for service leads to these

individuals being healthier and more highly educated than the general population. The army had minimum physical standards for enlistees to ensure that they were physically capable of meeting the demands of military service. The Selective Service System describes this need for standards in its special monograph on the physical examination of selective service registrants:

*The medical program of the Selective Service System during World War II was based on the need for selection of mentally alert and physically fit men who could make use of the tools of modern warfare to assure the final victory of the United Nations. Comcominant with this need, however, was the obligation of all who controlled this selection to protect from the rigorous demands of military life men who might be able to live useful and normal civilian lives but whose physical or mental defects would almost inevitably make them liabilities to the armed services under conditions of military life.*

Given this statement, it is obvious than many individuals facing childhood health shocks may not have passed the physical examination at their local induction station and would therefore not appear in the dataset of enlistee brothers. Determining exactly which individuals would be rejected by the army and therefore underrepresented in the enlistee sample is complicated. As manpower needs rose during the course of the war, the physical standards were relaxed. Between January 1942 and February 1946, there were 29 major changes to the physical standards, almost all relaxing those standards (see Table 19 in Selective Service System (1947) for a complete listing of these changes). Furthermore, as the war progressed the Army began to induct enlistees with certain correctible defects ranging from dental problems to illiteracy.

This variation in standards over time and how they were applied as well as issues with standards being interpreted differently at the various induction stations makes it difficult to draw a definitive cutoff for just how healthy and educated an individual needed to be to be accepted for military service (and therefore for the sample of enlistees used in this study). However, there are certain standards that remained fairly constant over time and certain aggregate statistics on the principal defects of selective service registrants that can provide a rough picture of the health of enlistees relative to rejected registrants.

Two sets of standards that remained constant over the war were the height and weight restrictions. Enlistees had to be between 60 to 78 inches tall and meet both minimum and maximum weight standards that varied with height. Minimum weights ranged from 105 pounds for a 60-inch-tall enlistee to 165 pounds for a 78-inch-tall enlistee. These standards suggest that individuals with severe stunting from childhood disease or individuals left extremely underweight from poor health will not be in the brothers sample. Note, however, that the lower height cutoff is extremely small. In data from the 1976 Integrated Health Interview Series, fewer than 1 percent of males born in 1920 were 60 inches tall or shorter. Fewer than 3 percent were 62 inches tall or shorter. Even if an individual experienced stunting of one to two inches from childhood disease, a very large effect, that individual would still be highly likely to make the height cutoff for military service.

While the height cutoff presents only minor concerns of sample selection, the rejection of registrants on the basis of other physical defects may be far more problematic. Chronic medical conditions leading to poor childhood health could follow the child into adulthood and lead to rejection as could physical defects that remained from a childhood disease that the individual overcame. Of the roughly 18

TABLE A1  
NUMBER OF REGISTRANTS REJECTED BY PRINCIPAL CAUSE  
FOR REJECTION

	Number of Registrants	Percentage of Total Number of Men Physically Examined
Total	4,828,000	27.30
Manifestly disqualifying defects	510,500	2.89
Mental disease	856,200	4.84
Mental deficiency	676,300	3.82
Physical defects	2,708,700	15.32
Within physical defects:		
Musculoskeletal	367,700	2.08
Cardiovascular	317,500	1.80
Hernia	260,000	1.47
Syphilis	254,800	1.44
Neurological	235,400	1.33
Eyes	234,300	1.32
Ears	189,700	1.07
Tuberculosis	129,900	0.73
Lungs	89,900	0.51
Underweight and overweight	69,600	0.39
Feet	69,200	0.39
Abdominal viscera	64,700	0.37
Kidney and urinary	53,300	0.30
Varicose veins	48,200	0.27
Genitalia	48,000	0.27
Endocrine	45,300	0.26
Teeth	36,200	0.20
Nonmedical	76,300	0.43

*Notes:* Defects found in less than 0.2 percent of all registrants are excluded from the table. Manifestly disqualifying defects include total deafness or blindness, both arms or legs missing, similarly serious defects, and being under treatment at a mental institution or for chronic or severe mental or physical disorders. Mental deficiency includes registrants disqualified for not meeting educational standards.

*Sources:* Numbers are taken from Table 27 of Selective Service System Special Monograph 15, "Physical Examination of Selective Service Registrants."

million Selective Service registrants between 18 and 37 years of age physically examined, nearly three million were rejected for physical defects. The number of registrants rejected by principal cause is shown in Table A1. Musculoskeletal defects and cardiovascular defects top the list of principal physical causes for rejection, causes that could easily be correlated with childhood health complications. Beyond these

causes, there was a wide range of other causes potentially related to childhood health shocks that led to rejection for hundreds of thousands of registrants.

The concern raised by the large numbers of rejections for a range of medical causes in Table A1 is that the registrants who did have severe childhood health issues did not pass the physical examination and therefore do not appear in the sample of brothers. Consequently, the analysis of childhood health impacts on educational investment pertains only to those childhood health shocks that were mild enough to leave the individual fit for military service. It is important to note, however, that given the manpower needs of the military in World War II, enlistees were not drawn from only those individuals with perfect health. Table A2 shows the prevalence of physical defects by the result of the examination. Physical defects were clearly not limited to those disqualified from service. Even among registrants accepted for general service with no required remediation of defects, the rate of physical defects is remarkably high: more than 40 percent of registrants accepted for general service had at least one physical defect. Enlistees deemed healthy enough for service still exhibited the musculoskeletal, cardiovascular, and other medical problems potentially indicative of poor health in childhood. While individuals with extremely severe childhood health shocks will not be present in the brothers sample, the figures in Table A2 nonetheless suggest that there will still be potentially large variation in health between enlisted brothers.

The standards for military service also present issues for the ability of the brothers sample to capture variation in schooling. Registrants had to meet minimum educational requirements to enlist. As with the physical standards, these educational standards varied over the course of the war as the need for manpower rose. The initial requirement was that enlistees needed the ability to read and write English on a fourth grade level. However, beginning in August 1942, registrants who were educationally deficient were accepted on a quota basis. These quota restrictions were lifted in June 1943 when the War Department introduced special training for illiterate enlistees. Table A2 demonstrates that these requirements were binding. Nearly 4 percent of registrants were found to be mentally deficient. The primary reason for mental deficiency was illiteracy; roughly 70 percent of white registrants and 90 percent of black registrants rejected on the basis of a mental deficiency had an educational deficiency (Selective Service System 1947, Table 51 and Table 52). Clearly the sample of brothers will be underrepresentative of individuals with very low levels of education. Just as the health requirements will lead to regression results that are specific to individuals with sufficiently mild long-term effects of childhood health shocks to be fit for service, the educational requirement leads to results specific to families that were investing a sufficient amount in both brothers, including the less healthy brother, for both to achieve literacy. Both the health and educational restrictions imposed by the Selective Service System imply that this study's results fail to capture the largest health shocks and the biggest gaps in educational investment across siblings. It is very possible that if these cases could be included in the brother sample the impact of health on educational investments would be even larger than the results reported in the article.

TABLE A2  
 PERCENTAGE OF REGISTRANTS WITH A PRINCIPAL DEFECT BY RESULT OF  
 EXAMINATION, NOVEMBER 1940-DECEMBER 1943

	<u>Result of examination</u>				
	Total	General Service	Available for Military Service after Correction of Defects	Available for Limited Service	Disqualifi ed from Service
Total	100	100	100	100	100
Eyes	8.7	6.0	1.6	31.7	5.4
Ears	2.5	1.0	0.9	2.5	5.5
Teeth	6.9	6.7	5.8	18.9	3.0
Mouth and gums	1.0	1.3	1.4	0.6	0.4
Nose	1.7	2.3	1.1	1.6	0.8
Throat	0.9	1.5	0.2	0.2	0.1
Lungs	0.9	0.3	2.2	0.9	2.1
Tuberculosis	1.3	0.2	4.7	0.2	3.7
Cardiovascular	4.4	1.3	1.9	2.3	11.3
Blood and blood-forming	0.1	<0.05	0.1	<0.05	0.1
Hernia	3.1	1.2	34.4	6.3	4.1
Kidney and urinary	0.6	0.2	3.0	0.6	1.2
Abdominal viscera	1.0	0.7	0.7	0.5	1.7
Genitalia	2.0	2.5	7.3	1.9	0.7
Syphilis	2.5	0.7	2.7	6.4	4.7
Gonorrhea and other venereal	0.4	0.2	5.9	0.9	0.2
Skin	1.2	1.6	1.4	0.6	0.6
Hemorrhoids	0.6	0.7	2.2	0.5	0.3
Varicose veins	0.9	0.7	2.1	0.9	1.0
Mental deficiency	4.3	0.5	0.8	0.3	13.3
Mental disease	4.7	0.4	1.7	0.9	15.0
Neurological	2.2	0.1	0.9	1.1	6.6
Musculoskeletal	5.6	2.7	5.5	9.6	9.8
Feet	4.1	5.4	0.2	4	1.9
Endocrine	0.7	0.2	0.4	0.6	1.6
Neoplasms	0.5	0.5	3.0	0.4	0.5
Infectious and parasitic	<0.05	<0.05	0.8	<0.05	0.1
Other medical	3.5	3.6	7	5.1	2.6
Non-medical	0.6	0.1	0.1	<0.05	1.6
No defects, and not stated	33.1	57.4	less than .05	0.5	0.1

Sources: Figures are taken from Table 79 of Selective Service System Special Monograph 15,  
 "Physical Examination of Selective Service Registrants."

*Selection Issues Related to Matching Enlistment Records to the Federal Census*

The process of linking brothers in the enlistment records using the federal census creates a different set of sample selection issues. In particular, the restrictions that potential brothers share the same state of birth and the same county of residence bias the sample in favor of less geographically mobile families and individuals. These restrictions lead to two potential sources of sample selection bias. First, the restriction on potential brothers sharing the same birth state eliminates a potentially large number of geographically mobile families. This will be problematic if geographically mobile families differ in the way they reallocate resources in response to health shocks. Second, the restriction on brothers residing in the same county will cause geographically mobile individuals to drop out of the sample. Of particular concern is that a negative health shock may make one brother less geographically mobile than his healthier sibling, causing the sample to be underrepresentative of precisely those pairs of brothers for which the relationship between health and educational attainment within the family is strongest. This section will discuss the family and individual characteristics correlated with geographical mobility and the consequences of these correlations for the estimated relationship between health and education.

One simple approach to assessing the impact of restricting the sample to brothers born in the same state is to compare families with all sons born in the same state to families with sons born in different states using the 1930 federal census, a time period when most enlistees were likely to be living with their siblings in their parents' household. Tables A3 and A4 present summary statistics for brothers and household heads for all households with at least two sons in the IPUMS 5 percent sample of the 1930 federal census. This sample is divided on the basis of whether all sons in the household share the same state of birth. From Table A3, it is clear that sets of brothers sharing the same birth state differ very little on average from brothers with different birth states in terms of education (at least in terms of the minimal education variables available in the 1930 census). There is no significant difference in the likelihood of being in school between the two groups and while the difference in literacy rates between the groups is statistically significant, it is quite small in magnitude; literacy among brothers with different birth states is 1 percentage point higher than among brothers born in the same state. Turning to the parents of these brothers in Table A4, household heads of families with brothers born in different states tend to have higher income jobs, are less likely to be farmers, are more likely to live in urban areas, and have larger families than household heads of families with all sons born in the same state.

These characteristics may all be correlated with how families make educational investment decisions and the types of schooling and health environments faced by the family. However, as long as they affect educational investments in the same way for all brothers in the family, the family fixed effects will adequately control for these average differences in characteristics between the two groups. The main threat to identification is if any of these characteristics affect the likelihood of families varying the distribution of resources among children. For example, farmers may be more likely to invest little in the formal education of the son taking over the family farm and more in the son expected to move into a different career. Parents who are not farmers may invest more equally in all sons given that they are all preparing for non-agricultural jobs. Health shocks may also have differential impacts across these families. For the farmers, a health shock that leads to permanent physical disability may lead to parents switching from preparing a child for a physically demanding agricultural job to preparing him for a white collar job that requires greater education. This may then lead

TABLE A3  
SUMMARY STATISTICS FOR BROTHERS BY GEOGRAPHIC  
MOBILITY, 1930

	Families with All Brothers Born in the Same State			Families with Brothers Born in Different States		
	Mean	Overall Standard Deviation	Within Standard Deviation	Mean	Overall Standard Deviation	Within Standard Deviation
<u>All sets of brothers</u>						
Age	12.04	7.81	3.87	12.55	7.91	4.64
Attending school (yes=1)	0.50	0.50	0.36	0.51	0.50	0.39
Literate (yes=1)	0.57	0.50	0.32	0.60	0.49	0.35
Number of observations		805,168			126,162	
<u>Only 14- to 17-year-old brothers</u>						
Age	15.46	1.11	0.57	15.51	1.17	1.03
Attending school (yes=1)	0.71	0.45	0.16	0.73	0.45	0.30
Literate (yes=1)	0.98	0.15	0.04	0.99	0.10	0.05
Number of observations		155,643			3,047	

*Notes:* Within standard deviation is calculated by subtracting the family mean of the variable and then calculating the standard deviation of the resulting demeaned variable. The sample is restricted to males listed as the child of the head of the household with at least one brother present in the household.

*Sources:* IPUMS 5 percent sample of the 1930 federal census.

TABLE A4  
HOUSEHOLD HEAD CHARACTERISTICS BY GEOGRAPHICAL MOBILITY,  
1930

	All Brothers Born in the Same State	Brothers Born in Different States
Age	44.25 (10.53)	44.84 (9.74)
Farmer (1=yes)	0.33 (0.47)	0.21 (0.41)
Living in an urban area (1=yes)	0.48 (0.50)	0.59 (0.49)
Literate (1=yes)	0.92 (0.26)	0.92 (0.26)
Employed (1=yes)	0.87 (0.34)	0.84 (0.37)
Income (in hundred of 1950 dollars)	21.21 (11.58)	22.38 (11.85)
Number of children in household	4.07 (1.78)	4.33 (1.83)
Number of observations	302,082	43,168

*Notes:* Standard deviations given in parentheses. The sample includes all household heads with two or more sons present in the household.

*Sources:* IPUMS 5 percent sample of the 1930 federal census.



to a larger response to a health shock in terms of shifts in educational investments relative to an urban family that was already preparing all children for white collar jobs.

To assess whether variation in educational investments across brothers differ with the geographic mobility of the household the family mean can be subtracted from the education variables to create a new set of demeaned education variables. The variation in these demeaned variables captures the variation within rather than across households. The standard deviations of these demeaned variables are reported in Table A3 as the “within standard deviation.” Focusing on those sons for whom variation in school attendance is likely to be most pronounced, those of high school age, Table A3 demonstrates that there is greater within family variation in school attendance for geographically mobile families suggesting that these families may be more likely to vary their resource allocations across children. These findings suggest that by including only brothers born in the same state, the sample may be underrepresentative of the families most likely to vary educational investments across children. The sample will also be underrepresentative of families with large within-family variation in the health environments children were exposed to. If the relationship between health and education follows the same pattern found in the article of increased educational investments in healthy children relative to less healthy children, the estimated results excluding the most mobile families will likely underestimate the true relationship between childhood health shocks and educational attainment. This underestimation is the product of both a greater attenuation bias from excluding the families for which variation in height is more strongly correlated with childhood health shocks and a sample selection bias from excluding families most willing to vary their educational investments across children.

The second issue of geographical mobility is the restriction that brothers must reside in the same county at the time of enlistment. This will eliminate both sets of brothers for which both brothers are very geographically mobile and sets of brothers for which one brother is geographically mobile and the other is not. To consider the ways this biases the sample, Table A5 shows the characteristics of males in their 20s in the IPUMS 1 percent sample of the 1940 federal census. These individuals are divided into two groups on the basis of whether they lived in a different state five years earlier.<sup>1</sup> More geographically mobile individuals were both more educated in terms of years of schooling and more successful in terms of occupational outcomes and annual income. The sample of brothers used in the article will therefore be underrepresentative of more educated, higher-income individuals. This is particularly problematic for pairs of brothers in which one brother is geographically mobile. These pairs will drop out of the sample because of the mobile brother and will have larger gaps in education and income between brothers than the average pair of brothers remaining in the sample. Losing these pairs of brothers will at best lead to less precise estimates of the relationship between height and education due to the loss of substantial variation in the data. More problematic is the possibility of a nonlinear relationship between height and education. Losing those pairs of brothers with the largest gaps in height and education would lead to over- or underestimates of the marginal effect of height on education if that marginal effect is either decreasing or increasing with height, respectively.

---

<sup>1</sup> While it would be ideal to identify individuals that have moved across counties, the closest the 1940 census comes to identifying this information is including whether the individual lived in the same house as he did five years earlier and whether he lived in the same state as he did five years earlier. Given that many of these individuals in their 20s may have been moving out of their parents' homes in the previous five years, the information on whether they live in a different house is not a particularly useful proxy for geographic mobility. Whether the individual moved across states is therefore the best measure for geographic mobility of the sort relevant here.

TABLE A5  
INDIVIDUAL CHARACTERISTICS FOR 20- TO 29-YEAR-OLD MALES BY  
GEOGRAPHIC MOBILITY, 1940

	Did Not Move across States in the Previous Five Years	Moved across States in the Previous Five Years
Years of schooling	9.66 (3.30)	10.81 (3.38)
Employed (1=yes)	0.81 (0.39)	0.84 (0.37)
Weeks worked in past year	41.79 (13.86)	41.83 (13.50)
Hours worked in past week	45.36 (12.94)	46.02 (12.93)
Wage and salary income in previous year (in 1940 dollars)	597.10 (608.86)	789.80 (703.88)
Median income for occupation (in hundreds of 1950 dollars)	19.74 (10.95)	21.39 (12.15)
Number of observations	100,646	9,592

*Notes:* Standard deviations given in parentheses. Median income is based on the median income in 1950 for all Americans with that occupation.

*Sources:* IPUMS 1 percent sample of the 1940 federal census.

### *Public Health Reports*

The Public Health Reports are a weekly publication of the United States Public Health Service. They have been published since 1887. The typical weekly report contains articles on current public health issues and research findings and then a section on the prevalence of disease. The prevalence of disease section gives the number of cases and deaths reported for various diseases by states and cities in the previous week.

In the early 1920s, the Public Health Reports would include an annual summary of the prevalence of disease in the previous year. One issue presented the annual summary for cities with a population over 100,000 and a second issue presented the annual summary for cities with a population between 10,000 and 100,000. The morbidity and mortality data for cities used in the article come from the 1926 annual summary, the last summary published for cities with populations greater than 100,000 (although weekly reports continued to be published). The small city data is of questionable quality, with many of the cities failing to report information for several of the diseases and warnings from the Public Health Service that reporting standards for the cities were changing a great deal over the period of interest.

The annual summary contains the total number of cases and deaths in the previous year for a variety of diseases including anthrax, cerebrospinal fever, chicken pox, dengue fever, diphtheria, influenza, lethargic encephalitis, malaria, measles, mumps, pellagra, pneumonia, poliomyelitis, rabies in animals, rabies in man, scarlet fever, septic sore throat, smallpox, tuberculosis, typhoid fever, typhus fever, and whooping

cough. Cases and deaths are reported in both absolute numbers and in per capita terms. In addition to the number of cases in the previous year, the summary includes what the Public Health Service calls the “estimated expectancy.” This figure is the expected number of cases in a non-epidemic year and in most cases is calculated as the median number of annual cases reported between 1918 and 1924, inclusive. If epidemics occurred, those years are excluded and the estimated expectancy is calculated as the mean of the number of cases reported in non-epidemic years. The number of years of data used for each calculation is reported in the tables. No estimated expectancies were given for anthrax, influenza, lethargic encephalitis, malaria, pellagra, pneumonia, rabies, tuberculosis, or typhus fever. For these diseases, I use the number of reported cases and deaths in 1925 in place of the missing estimated expectancies.

For several cities, the public health reports did not include a population estimate. In these cases, I have imputed the 1925 population by taking the average of the city populations reported in the 1920 and 1930 federal censuses. This was done for the following cities: Los Angeles (CA), Bridgeport (CT), Waterbury (CT), Atlanta (GA), Elizabeth (NJ), Akron (OH), Oklahoma City (OK), Portland (OR), Erie (PA), Houston (TX), Norfolk (VA), and Seattle (WA).

One major note of caution when using the Public Health Reports is that the ability to diagnose diseases and the efforts to report cases were changing over time. This makes it difficult to interpret changes over time in the number of cases, the number of deaths and in the ratio of deaths to cases. The Public Health Service included the following warning with each annual summary:

“In comparing the figures for 1925 with the estimated expectancy, averages, or with reports for preceding years, it should be borne in mind that for several years there has been a gradual improvement in the reporting of communicable diseases. An increase in the number of cases reported may be due to better reporting rather than to an increase in the number of cases occurring.”

### *1920 and 1930 Federal Censuses*

The 1920 and 1930 federal censuses are used to identify brothers and gather information on their families. The process of matching individuals to the 1920 and 1930 census is described in the section on the World War II enlistment records. The purpose of this section is to elaborate on the information available in the federal census, the differences between the censuses, and the limitations of the census data.

The forms for the 1920 and 1930 federal censuses are very similar. The 1930 census includes the following variables relevant to this study: full name, age, state or country of birth, occupation, industry, whether the household head owns or rents the residence, what the monthly rent or value of the home is, and relationship to head of household. The 1920 census includes all of these variables with the exception of the monthly rent or value of the home.

Once sets of brothers are found in the census, the number and ordering of siblings is recorded as is the information for the head of household. In nearly all cases, the head of household is the father of the brothers. In rare cases, the head of the household is a single mother, a grandparent, or another relative. In cases where one brother is listed as a son while the another is listed as a step-son, the brothers are dropped from the sample (the identification strategy depends in part on brothers having the same parents by birth).

To determine the income of the head of household, I match the listed occupation for the household head to the occupations from the 1950 federal census. This allows me to assign a 1950 occupational income score to the household head. The 1950 occupational income score is based on the median income in hundreds of 1950 dollars for each particular occupation. While the income distribution by occupation in 1950 is certainly different than that of 1920 or 1930, these 1950 occupational income scores offer the best income estimates available for the household heads in the sample. Reliance on the the 1950 occupational income scores does mean that the income variable may be a noisy proxy of actual household income.

More information on the construction of the occupational income scores and the occupational coding in the federal census can be found on the Integrated Public Use Microdata Series website ([usa.ipums.org](http://usa.ipums.org)). The site also contains information on the full set of variables available in the 1920 and 1930 federal censuses.

### *1880 Federal Census*

The 1880 federal census was unique for its collection of morbidity information. With the exception of the 1880 census, the federal census did not ask detailed questions about disability until 1970. While several decades removed from the period of interest in this article (a major concern given the substantial improvements in health at the start of the twentieth century), it offers the only opportunity to get age distributions of morbidity rates for several diseases from a large, nationally representative sample of the population. The age distributions of deaths by disease, published annually by the census bureau, do demonstrate that the age distributions of cases in 1880 are quite similar to the age distributions of deaths in the 1920s, helping minimize concerns about the applicability of the 1880 data to the enlistees. These figures from the annual mortality statistics are included in Table 2 of the article.

The census asked the following question: “Is the person (on the day of the enumerator’s visit) sick, or temporarily disabled, so as to be unable to attend to ordinary business or duties? If so, what is the sickness or disability?” The phrasing of the question appears to focus on work-related disabilities raising concerns that it would not apply to children and would therefore not be useful to study the incidence of childhood disease. However, it appears from the age distribution of individuals reporting an illness that the question was treated more generally.

Of those reporting an illness or disability, 10 percent were below the age of four and 25 percent were below the age of 12. Given the large percentage of illnesses reported for individuals far too young to work, it seems likely that a sizable percentage of individuals were interpreting the question as asking about any illnesses on the day of the census, not simply illnesses that were interfering with work. However, these percentages are still lower than what we would expect and suggest that the morbidity data is providing an underestimate of childhood morbidity rates relative to adult morbidity rates.

A second caution about the interpretation of the 1880 census morbidity data is that all of the illnesses are self-reported (or reported by parents). It is certainly possible that individuals are misdiagnosing their illnesses, exaggerating their illnesses, or even hiding their illnesses. All of these possibilities contribute additional noise to the morbidity data.

ADDITIONAL TABLES AND FIGURES

TABLE A6  
THE EFFECT OF MORTALITY RATES ON HEIGHT, CITY OR STATE  
MEAN HEIGHT AS DEPENDENT VARIABLE

	City Level Disease Data		State Level Disease Data	
	(1)	(2)	(3)	(4)
Mortality rate due to diseases	-0.025***	-0.008	-0.015***	-0.015***
targeting infants	(0.008)	(0.007)	(0.003)	(0.003)
Mortality rate due to diseases	0.004	0.003	0.005	0.008
targeting older children	(0.010)	(0.008)	(0.006)	(0.007)
Mortality rate due to diseases	0.029***	0.016***	-0.002***	-0.001**
targeting adults	(0.006)	(0.006)	(0.001)	(0.001)
<u>Region dummies:</u>				
Northeast	—	-0.636***	—	-0.141
	—	(0.099)	—	(0.107)
South	—	-0.154	—	0.040
	—	(0.128)	—	(0.071)
West	—	0.222*	—	0.103
	—	(0.116)	—	(0.068)
Constant	68.307***	68.428***	69.694***	69.465***
	(0.122)	(0.100)	(0.165)	(0.183)
Observations	64	64	47	47
R-squared	0.29	0.63	0.82	0.83

\* = Significant at the 10 percent level.

\*\* = Significant at the 5 percent level.

\*\*\* = Significant at the 1 percent level.

Notes: Robust standard errors in parentheses. Unit of observation is an individual city for columns (1) and (2) and an individual state for columns (3) and (4). Omitted region dummy is for the Midwest. All mortality rates are deaths per 100,000 people.

Sources: World War II Army enlistment records, reports of the Department of Public Health, and Grant Miller's database of state mortality rates.

TABLE A7  
OLS REGRESSION RESULTS FOR THE EFFECTS OF HEIGHT AND HOUSEHOLD  
CHARACTERISTICS ON EDUCATION, YEARS OF SECONDARY AND  
POST-SECONDARY EDUCATION AS DEPENDENT VARIABLE

	(1)	(2)	(3)	(4)
Height (inches)	0.074*** (0.010)	0.068*** (0.010)	0.075*** (0.010)	0.071*** (0.009)
Number of brothers	— —	−0.333*** (0.022)	−0.301*** (0.022)	−0.270*** (0.024)
Birth order among brothers	— —	0.097*** (0.025)	0.091*** (0.024)	0.093*** (0.024)
Ln(father's income)	— —	— —	— —	1.019*** (0.076)
Race and year fixed effects	no	no	yes	yes
Observations	8456	8198	8154	7459
R-squared	0.03	0.07	0.14	0.18

\* = Significant at the 10 percent level.

\*\* = Significant at the 5 percent level.

\*\*\* = Significant at the 1 percent level.

*Notes:* Robust standard errors clustered by childhood state of residence in parentheses. All regressions control for a quadratic in age. Only individuals with completed educational careers are included in the regression sample.

*Sources:* World War II Army enlistment records linked to the 1930 federal census.

TABLE A8  
LOGIT ESTIMATES OF THE EFFECTS OF HEIGHT AND HOUSEHOLD  
CHARACTERISTICS ON EDUCATION OUTCOMES

	Attended at Least One Year of High School (yes=1)	High School Graduate (yes=1)	Attended at Least One Year of College (yes=1)
Dependent Variable:	(1)	(2)	(3)
Height (inches)	0.059*** (0.013)	0.084*** (0.011)	0.104*** (0.013)
Number of brothers	−0.243*** (0.035)	−0.292*** (0.025)	−0.446*** (0.064)
Birth order among brothers	0.069** (0.029)	0.114*** (0.025)	0.121 (0.082)
Ln(father's income)	1.185*** (0.098)	0.877*** (0.094)	1.329*** (0.130)
Observations	7451	7455	5951

\* = Significant at the 10 percent level.

\*\* = Significant at the 5 percent level.

\*\*\* = Significant at the 1 percent level.

*Notes:* Robust standard errors clustered by childhood state of residence in parentheses. All regressions control for race, birth state, and a quadratic in age. Only individuals with completed educational careers are included in the regression sample.

*Sources:* World War II Army enlistment records linked to the 1930 federal census.

TABLE A9  
ESTIMATES OF THE EFFECT OF HEIGHT ON EDUCATIONAL ATTAINMENT  
WITH FAMILY FIXED EFFECTS

Dependent variable:	Years of Secondary and Post-Secondary Education	Attended at Least One Year of High School (yes=1)	High School Graduate (yes=1)	Attended at Least One Year of College (yes=1)
	(1)	(2)	(3)	(4)
Height (inches)	0.029*** (0.010)	-0.006 (0.021)	0.073*** (0.026)	0.063** (0.031)
Birth order among brothers	0.058 (0.043)	0.053 (0.096)	0.220*** (0.083)	0.173 (0.218)
Observations	8225	1724	1913	569
R-squared	0.01	0.02	0.03	0.11

\* = Significant at the 10 percent level.

\*\* = Significant at the 5 percent level.

\*\*\* = Significant at the 1 percent level.

*Notes:* Robust standard errors clustered by family in column (1) and childhood state of residence in columns (2), (3), and (4) in parentheses. All regressions control for a quadratic in age. Only individuals with completed educational careers are included in the regression sample. Column (1) is an OLS regression. Columns (2), (3) and (4) are conditional logit regressions.

*Sources:* World War II Army enlistment records linked to the 1930 federal census.

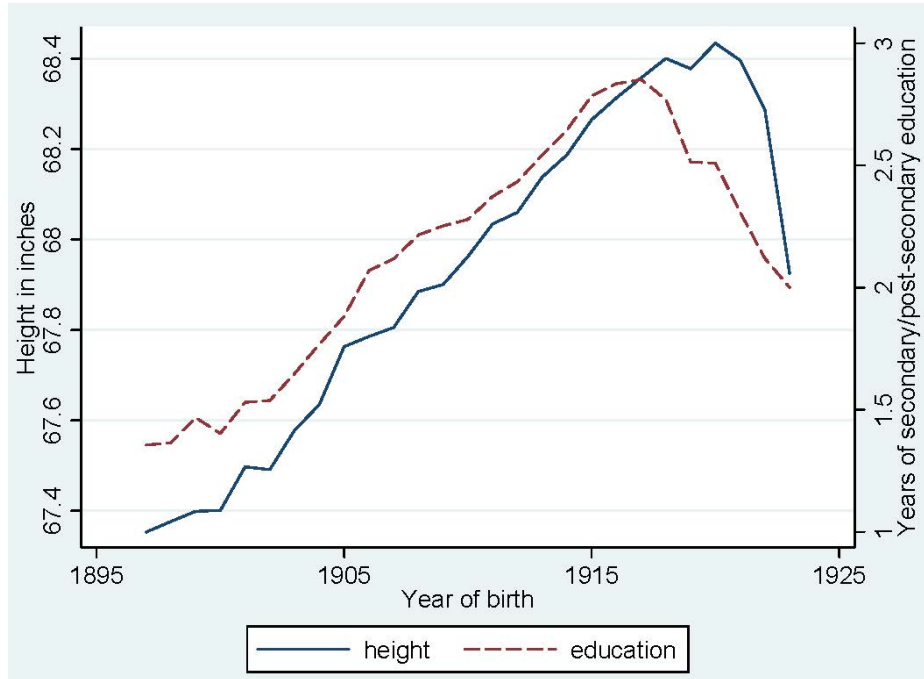


FIGURE A1  
 MEAN HEIGHT AND EDUCATIONAL ATTAINMENT BY  
 COHORT FOR PRIVATES WITH COMPLETED EDUCATIONAL  
 CAREERS, 1897–1923.

*Notes:* The dropoff in height and educational attainment for the youngest cohorts is a product of conditioning on completed educational careers: the youngest enlistees could only have completed their education if they received relatively few years of schooling.

*Sources:* World War II Army enlistment records.



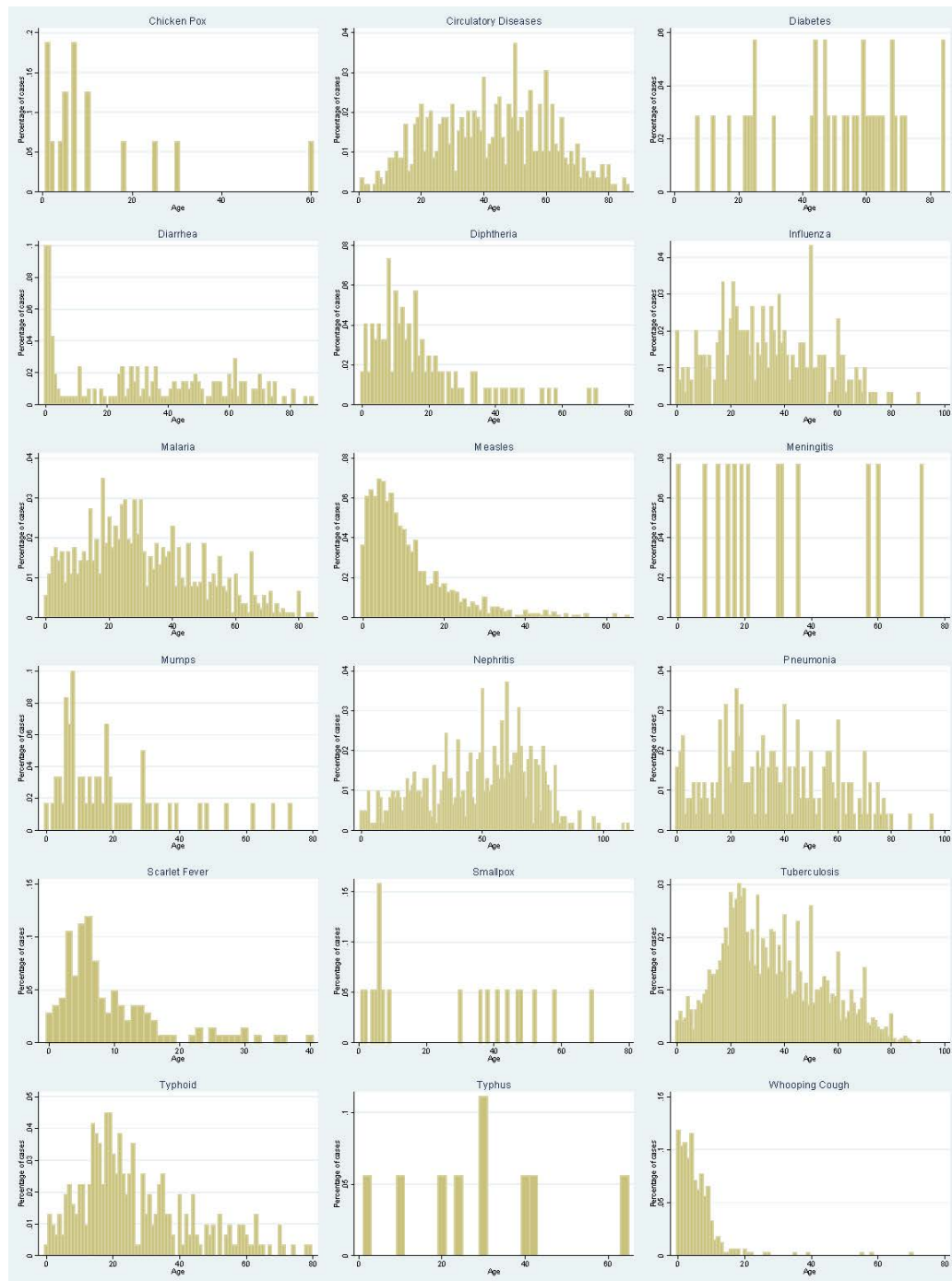


FIGURE A2  
AGE DISTRIBUTIONS FOR CASES OF MAJOR DISEASES AS  
REPORTED IN THE 1880 FEDERAL CENSUS

Sources: IPUMS 1 percent sample of the 1880 federal census.

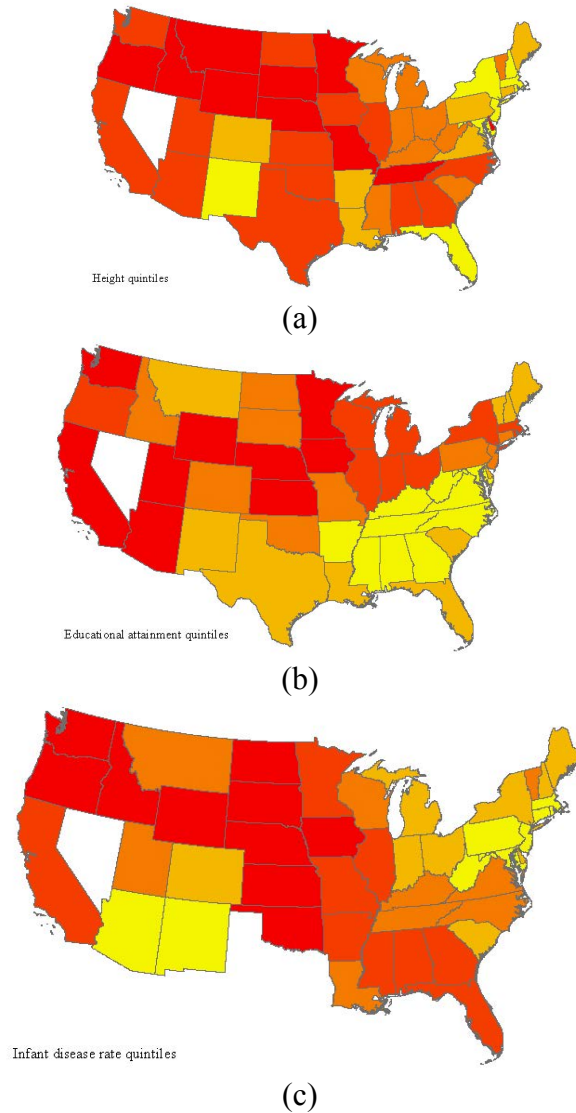


FIGURE A3  
 AVERAGE HEIGHT (A), EDUCATION (B) AND MORTALITY  
 RATE DUE TO DISEASES TARGETING INFANTS (C) BY STATE

*Notes:* Colors correspond to quintiles of each variable's distribution. Yellow (lightest shade) corresponds to the lowest height and education quintiles and the highest infant mortality quintile. Red (darkest shade) corresponds to the highest height and education quintiles and the lowest infant mortality quintile.

*Sources:* World War II Army enlistment records and Grant Miller's database of state mortality rates.

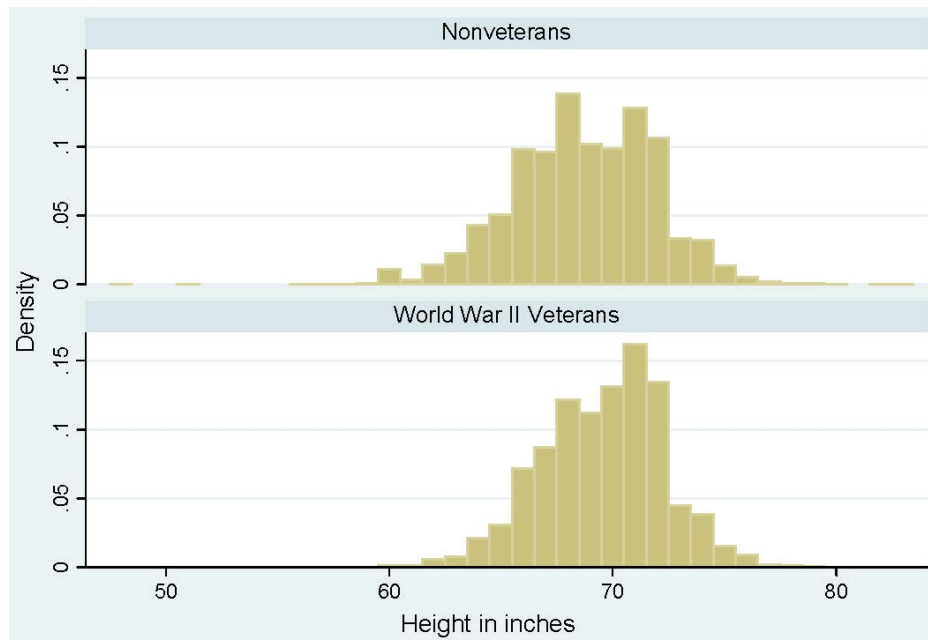


FIGURE A4  
DISTRIBUTION OF MALE HEIGHTS FOR WWII VETERANS AND  
CIVILIANS IN THE 1976 INTEGRATED HEALTH INTERVIEW  
SERIES.

*Notes:* Civilian observations are weighted to match the age distribution of veterans.

*Sources:* Integrated Health Interview Series data.

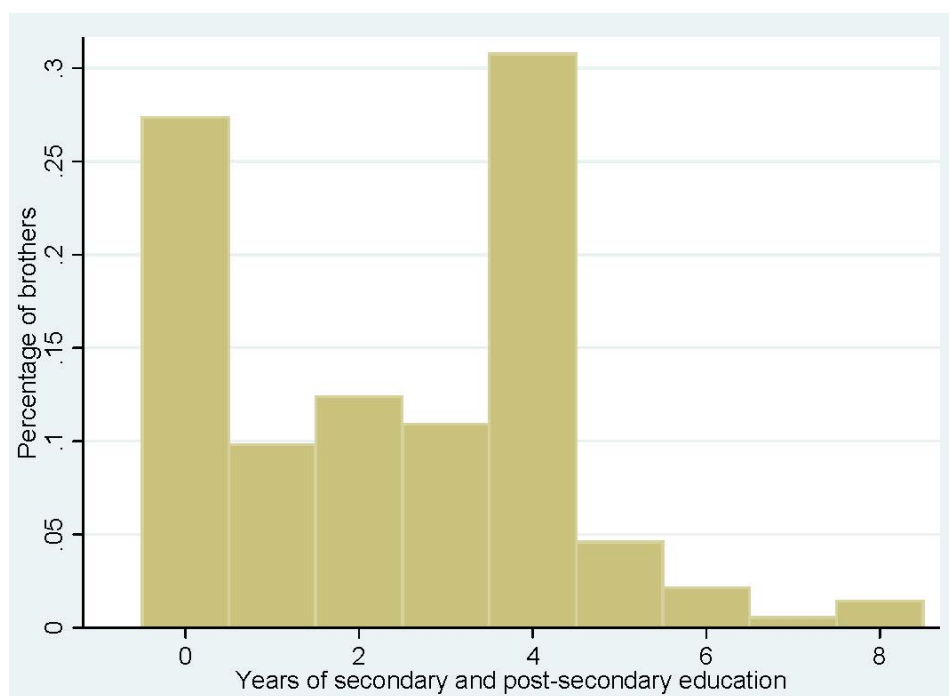


FIGURE A5  
DISTRIBUTION OF EDUCATIONAL ATTAINMENT FOR ALL  
INDIVIDUALS WITH COMPLETED EDUCATIONAL CAREERS  
IN THE SAMPLE OF BROTHERS

Sources: World War II Army enlistment records linked to the 1930 federal census.