Economics and Human Biology 34 (2019) 274–285

Contents lists available at ScienceDirect

Economics and Human Biology

journal homepage: www.elsevier.com/locate/ehb

Does adult height predict later mortality?: Comparative evidence from the *Early Indicators* samples in the United States

Sven E. Wilson

Brigham Young University, United States

ARTICLE INFO

Article history:
Received 3 September 2018
Received in revised form 11 May 2019
Accepted 24 May 2019
Available online 8 June 2019

Keywords:
U.S. mortality
Anthropometric history

ABSTRACT

In this paper, I supplement widely used demographic data on white veterans of the Union Army with large and newly collected data on blacks and urban white veterans to explore the question of whether adult height predicts late-life mortality at the individual level. The data are partitioned into four demographic groups based on individual characteristics at the time of enlistment: white veterans enlisting in rural areas, mid-size cities, and large cities, and African-American veterans of the U.S. Colored Troops (USCT). Across the three groups of white veterans, mean height is positively associated with life expectancy at age 60, while both mean height and life expectancy for black veterans are very close to levels measured among the highly urbanized white veterans. I examine whether these group-level differences are robust to individual-level analysis by estimating two types of models, separately for each group: 1) 10-year mortality at age 60 using a linear probability model with company-level fixed effects and 2) a Cox proportional hazard that tracks veterans from age 60 to death. For rural whites, I find a significant U-shaped relationship between height and 10-year mortality, with both the short and the tall at significantly higher risk of death. This pattern becomes more pronounced when excluding younger recruits (under aged 24) from the analysis. But this relationship does not extend to urban whites or to blacks, where no significant height effects are found, and in which the height-mortality relationship among the highest mortality groups (whites from the largest cities and blacks) appears to be a generally positive one. Overall, the robust positive relationship between height and life expectancy at the group level does not exist at the individual level.

1. Introduction

Fogel (1986) was the first to document that the long-term trend in height mirrored trends in mortality.[1] In brief, both adult height and life expectancy declined during the mid to late 19th century before increasing steadily throughout the 20th century. This research and many subsequent studies are important because they demonstrate that height, which is primarily determined by genetics, is also a valuable measure of population health.

The association between height and mortality can also be seen cross-sectionally by comparing mean height and mortality across populations and population sub-groups. For example, a few of the estimates obtained later in this paper are illustrated in Fig. 1 below. In this figure, I compare four distinct demographic groups of Union Army veterans at the turn of the 20th century (differentiated by race and level of urbanization at enlistment) by plotting life expectancy at age 60 against adult height, which was obtained from the enlistment records of the recruits.

The relatively tight linear fit of these four points strongly suggests that the connection between these important population health measures—adult height and life expectancy—is shaped by the social forces that differentiate these demographic groups from one another. As Komlos and Lauderdale (2007) note, “there is widespread agreement that nutritional intake, the incidence of diseases, and the availability of medical services have a major impact on human size” (295), and cities in 19th century America could be very unhealthy places, rife with low quality food, a host of infectious diseases, and high population density. Cities also had a high proportion of immigrants and large pockets of low-skilled workers. Even though black veterans were primarily raised in rural areas, most grew up as slaves and they attained, on average, a stature and life expectancy very similar to white Americans from the largest cities. Of course, other social variables not reflected in Fig. 1, such as ethnicity, income, and social class, also affect population health.

But the differences present across demographic groups of the type shown in Fig. 1 do not necessarily mean that height influences

E-mail address: sven@byu.edu (S.E. Wilson).

[1] See Costa (2015) for a recent update of the Fogel data.

https://doi.org/10.1016/j.ehb.2019.05.004
1570-677X/

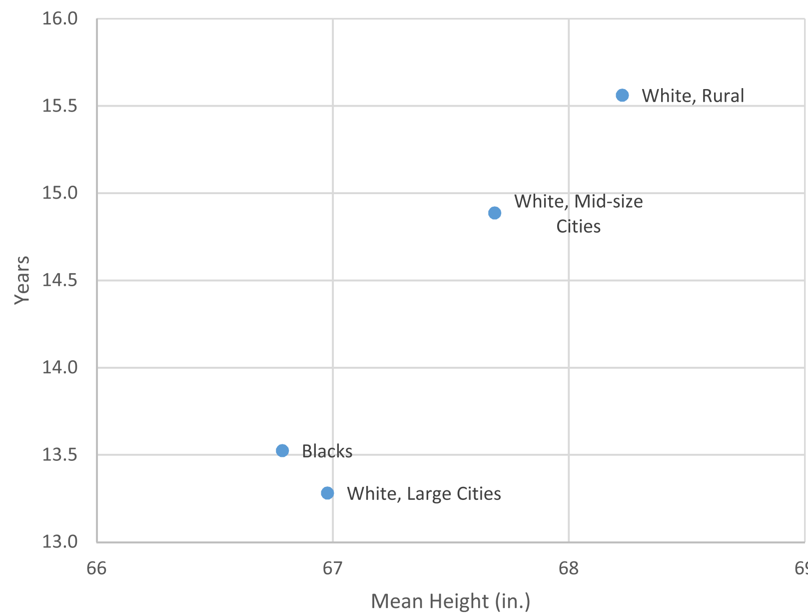

Fig. 1. Height and Life Expectancy at Age 60 across U.S. Demographic Groups, 1895-1904.
Notes: Data and estimation methods described in text below. Sample includes only those individuals with known death dates who enlisted between the ages of 20–39 and turned 60 between 1895-1905."Rural" means an enlistment city of less than 10,000 in 1860; "Mid-size" cities have 1860 population between 10,000 and 100,000; "Large" cities are the five largest cities in 1860: New York, Philadelphia, Baltimore, Boston, and Chicago. Height measured at enlistment.

mortality at the individual level. This paper, therefore, focuses on one aspect of the height-mortality relationship: after controlling for demographic and socioeconomic variables, does adult height predict mortality in later life? In looking at mortality after age 60, I find a U-shaped relationship among rural, white veterans, with both the shortest and the tallest individuals experiencing higher mortality, as some other historical research has shown. However, I find no such relationship between height and mortality among urban white veterans or among African-American veterans. Indeed, the estimated relationship among the higher mortality groups (whites in the largest cities and blacks) shows a generally positive (but not statistically significant) relationship between height and mortality

Even small effects of height on mortality should be detectable if datasets are large enough. Similarly, robust evidence would show that the height-mortality relationship exists across different subgroups of the population, as well as across populations. In this paper, I use the completed *Early Indicators* collections of Union Army veterans, which consist of nearly 40,000 whites, over 21,000 blacks from the United States Colored Troops (USCT), and an oversample of over 12,000 white, urban companies. Combined together, this collection of over 73,000 soldiers provides the largest data source that has, to my knowledge, ever been used for historical anthropometric analysis of mortality in the United States.[2]

2. Background

2.1. The importance of height in historical demography

In the 1970s, scholars became interested in the long-term decline in mortality over the past three centuries. Historical data on the determinants of mortality was very thin, but researchers began to see the utility of using records on height, mostly from military sources, to unpack the reasons for this decline. Fogel and Engerman (1974) had previously used data on heights to investigate the health of American slaves, and this work led Fogel and his collaborators (Fogel et al., 1978) to focus on other available data on heights that was available, mostly from military sources, in the United States and Europe.

With these newly developed datasets, researchers were able to see clear differences in mean height across populations and cohorts, but what could account for those differences?

Mean height reflects the net nutritional status of a population, andnet intake is a function of food consumed and energy expended. Nutritional status can also be undermined by the presence of both epidemic and endemic diseases. Nutritional deficits, especially when they occur during key growth periods, can lead to significant stunting. A variety of economic and social variables, such as real income, food prices, population growth, urbanization, and access to medical care, underlie the net nutritional status—and, hence, mean height—of a population (Komlos and Lauderdale, 2007).

The historical record suggests a natural relationship between height and mortality. It is not hard to see that the long-term trends tend to move in tandem, a pattern first brought to light by Fogel (1986). Fogel's data, updated and summarized by Costa (2015), show that life expectancy and height were both increasing gradually over the 1700s. However, from the mid 1700s to about 1830, height among white males was relatively stable, but it than began a gradual decline until almost the end of the century, falling about 3 cm (Costa, 2015). Similarly, male life expectancy at age 10 fell from 55.4 in 1797 to 47.8 in 1857, which was then reflected in the decline in height starting in the 1830s. When this decline in height and increase in mortality in the mid-19th century became apparent, it puzzled economic historians because this was a period characterized by rising national income. Given the robust relationship between income and health, why was height declining and mortality increasing even as income was rising?

This puzzle of declining health in a period of strong economic growth came to be known as "the "antebellum puzzle," and understanding the paradox became a primary drive of

[2] These sample sizes refer to the complete data collection, not to the analytical samples used in this analysis, which are considerably smaller because of necessary sample restrictions. Sample selection issues associated with the *Early Indicators* data (discussed below), severely constrain what methods can be employed without introducing survival bias into the estimation.

anthropometric research, especially among economic historians. Komlos (2012) recently summarized the debate (in which he was an active participant). According to his summary, the debate was divided between Komlos (and his collaborators), who argued that declining nutrition was the primary culprit behind declining heights, as outlined first in Komlos (1987), and Fogel (and his collaborators), who emphasized disease and other environmental factors in addition to nutrition. Then, in 2011, Fogel and co-authors (Floud et al., 2011) re-estimated the nutritional intake of the US population, coming very close to Komlos's 1987 findings. Though they noted that "migration and urbanization led to sanitary problems and accelerated disease contagions," they concluded, consistent with Komlos, that "the increase in agricultural productivity did not keep up with the rapid growth of the population and its food demands" (p. 298). A central part of the nutrition-based account advocated by Komlos was that some subpopulations did not experience a decline in height, while others did. A story based on access to food is easier to fit to these facts than one based on disease (though disease effects are clearly not uniform across the population).

Following the Civil War, the rapid urbanization of the population brought a variety of negative health effects. Wilson (2003), for instance, shows high and increasing levels of chronic respiratory disease at the end of the 19th century, likely due to the pollution associated with urbanization and industrialization. These negative effects of growth and urbanization continued to drag down mean height in the U.S. in the latter half of the 19th century (Costa, 2015). Eventually, however, the positive effects of explosive economic growth produced enough nutritional resources to offset the negative effects of urbanization and significant levels of in-migration. The 20th Century saw, in the U.S., a significant increase in mean height of almost 8 cm (170–177.8) and a sustained increase in life expectancy.

2.2. Height and mortality in modern populations

The importance of nutrition's impact on height remains salient today. Grasgruber et al. (2016), for instance, show that the modern variance in mean heights across countries can be mostly explained by differences in diet and nutrition. At the country level, over 80% of the variance in height is explained by simple nutritional variables (principally the quantity and source of proteins in the diet[3]); an enormous 20 cm in mean height separates Cambodia at the bottom of the distribution from Netherlands at the top! Importantly, only a small amount of additional variance can be explained by country-level socioeconomic variables.

Differences in height across demographic groups persist in the modern US, including by race. Interestingly, even though blacks are disadvantaged with respect to many common population health measures, such as life expectancy, blacks and whites do not differ much in height. The mean height among white males aged 20–39 is 178.2, while mean height for blacks of the same age is 177.4, a difference of only 0.8 cm. However, Hispanic men in that age group are 5.8 cm shorter than whites, and Asians are 5.9 cm shorter. The gap appears to be narrowing, though; at ages 40–59, Hispanics are 6.7 cm shorter than whites, and Asians are 7.3 cm shorter. These differences may be related somewhat to differences in place of birth due to immigration, but they are likely mostly due to differences in diet.

In a review of the determinants of height, Batty et al. (2009) argue that even though over 80% of the individual variation in height can be explained by genetic factors, environmental factors are clearly related to height, as evidenced by the "stepwise secular increase in height across a multitude of populations beginning in the late 19th century." (138) Such widespread increases include countries such as Finland (Silventoinen et al., 2001) which had almost no in-migration until recently, suggesting that immigration (and the associated mixing of genetic groups) is not likely related to the increase in height over time. A variety of socioeconomic variables affect net nutritional status. Following Batty and Leon (2002), studies related to socioeconomic status can be divided into group-level and individual-level analyses. A large number of group-level studies have been conducted, such as the work by Komlos and Kriwy (2002), which compared residents of East and West Germany. Individual-level studies of socioeconomic status report effects of occupation (Davey Smith et al., 2000), income (Steckel, 1995a,b), and educational attainment (Davey Smith et al., 1998b).

There is also an extensive international research on individual-level analysis of the height-mortality relationship. These results, however, are not uniform and include many null effects. Jousilahti et al. (2000) in Finland and Song et al. (2003) in South Korea, both impose a linear specification and find a significantly negative (but quite small) relationship between height and mortality. Sawada et al. (2017) find that in Europe height is positively associated with cancer mortality but inversely associated with cardiovascular mortality. In Switzerland, Rohrmaann et al. (2017) find a slightly negative (but insignificant) relationship between height and mortality among men and find a positive relationship for cancer mortality among women. They also note that their research "does not support an inverse association of body height with all-cause mortality." In Japan, Ihira et al. (2018) find that height is inversely associated with cerebrovascular mortality in men and women and respiratory mortality among men only, but is positively associated with cancer mortality among men. In the modern U.S., Sohn (2016) find that the taller actually die earlier, which they attribute to the persistence of high rates of cancer. In sum, the conventional wisdom that height and mortality are inversely related is not uniformly supported by the evidence. Null effects are common (even with very large samples), and positive associations exist in some places, especially for cancer. Furthermore, results differ significantly by gender.

Even though mean height may reflect some aspect of population health, this does not mean that it is necessarily a good individual-level predictor of mortality. The individual-level studies of height and mortality find generally modest (if any) effects. In one sense these small effects are not surprising: since most of the variation in height at the individual level is due to genetic factors, relatively little variation remains to be explained by other variables. It could be that the effects of individual height on later-life mortality are weak or non-existent, or it could be that controlling for other variables correlated with height, such as socioeconomic status, race, or nativity, will eliminate estimated impact of height. It is also possible that height affects mortality over some ranges of the life-cycle, but not others, and it is even possible that the relationship between height and mortality is a positive one.

3. Data & methods

3.1. The *Early Indicators* collections

In 1991, Bob Fogel received funding to begin an extensive collection of data on the lives of Union Army soldiers (Costa et al., 2017). The project, called *Early Indicators of Later Work Levels,*

[3] Rice-based diets (with relatively low protein) in East Asia lead to much lower height than the diets high in plant-based proteins consumed in many North African and Near Eastern nations, which are lower still than the diets of Europe that have a higher proportion of animal-based protein. Even highly developed East Asian economies lag significantly behind Northern Europeans.

Disease, and Death (hereafter, *Early Indicators),* began with a sample of 331 Union Army infantry companies containing close to 40,000 white soldiers (but not their officers). This sample was large enough to be roughly representative of the male, adult population at the time of the Civil War. Collection began in 1992 and took several years to complete fully. Additional samples have been added to the collection over time. In 2002, the research team began collecting a sample of about 6000 African American troops from the USCT (the United States Colored Troops). In 2006, the team began collecting an oversample of 12,000 recruits who served in companies formed in five large cities (New York (including Brooklyn), Philadelphia, Baltimore, Boston, and Chicago), and in 2010, they added an additional 15,000 USCT troops to bring the number of black veterans in the collection to over 21,000. After years of cleaning and standardizing the data, the complete collections have recently been made available to researchers.

The nature of the administrative data, especially the changing laws regarding pensions, and the manner in which the records were collected affect critically how the data should be used to avoid selection bias (more on that topic below). The initial step was to collect the military service records for each recruit in the selected companies; the veterans were then linked to the records of the Pension Bureau housed in the National Archives in Washington, DC. Pension files contain all the applications, rulings, and supporting documents associated with the veterans' applications for disability pensions in the decades after the war, though many of the veterans died before becoming legally eligible for a pension. Over time, the pension grew considerably in scope and, eventually, almost all living veterans were deemed eligible. These records contain significant demographic and economic data over the veterans' lives, including the date, place and (sometimes) cause of death. They also contain the detailed "Surgeons' Certificates," which contain all the detailed notes from the medical exams that pension applicants were subject to as part of the application process. Finally, the veterans were linked to, first, the federal census schedules of 1900 and 1910 and, then, to the records of 1850 and 1860. Parts of the collection have also been linked to the 1870, 1880, and 1920 census records. Thus, the Union Army recruits (and their family members) are observed, where possible, from early life until their death, and now the project is trying to learn about their children and grandchildren (Costa et al., 2017).

3.2. Pension law and survival Bias

Though the data cover the complete lifespan of thousands of Union Army recruits, a large hole exists in the data that must be circumnavigated by researchers. The basic problem is this: thousands of veterans are discharged from service and are never seen again. They never apply for a pension and, as a result, have no death date and are less likely to be found in census records either in childhood (1850 and 1860) or in later life (1900 and 1910). A missing death date can exist for a variety of reasons, but the biggest reason is that the veteran never applied for a pension. Thus, the sub-sample of veterans who have death dates consists mostly of pensioners, and pension eligibility is highly related to survival time.

The pension began during the war in 1863, though only a small fraction of soldiers received pensions before the end of the war. Initially, the applicant had to demonstrate that he had a disability that limited his potential to perform manual labor; he had to verify his time of service (a 3-month minimum was required); and he had to provide evidence that his disability was related to his service in the war. Many wounded veterans received a pension following the war, but most veterans were not sufficiently wounded to qualify and, therefore, did not apply. However, in the decades following the war, political pressure mounted for pension reform, due in part to the influence of veteran groups such as the Grand Army of the Republic. This pressure created an informal liberalization where more and more veterans entered the system under auspices of "war-related" disabilities (Wilson, 2010). Eventually the political pressure resulted in a formal change to the law in 1890, which dropped the requirement that pensionable disability be war-related. Following the passage of the 1890 law, a flood of veterans (both white and black) applied for and received pension support.

This history has a profound impact on how the data needs to be used and what kind of questions can be asked. In most cases it is not possible to know whether a recruit's death date is missing because he died before liberalization or because he lived but never applied for a pension. Ignoring this fact can lead to serious survival bias, since the factors that affect pension eligibility can also affect survival. Probably the most important factor affecting enrollment in the pension is race, due to the discrimination that blacks faced in the pension system and, before that, during the war (Wilson, 2010). Thus blacks dying before 1890 are much less likely to have death dates than whites, which is one reason the models estimated later are applied separately for blacks. Fortunately, I know of no reason why our main variable of interest, height, should affect the probability of being pensioned other than through the pathway of health.

Another similar problem is that much of the data collected by the *Early Indicators* project is dependent on the existence of the pension records—including for data that occurred early in life. Demographic information is available from the census, but census information is less likely to be present if a pension was never applied for. The 1850 and 1860 census records can provide early-life data, but they were collected using information obtained from the pension and from the 1900 or 1910 census schedules. Thus, the existence of the census information is highly dependent on survival. To avoid introducing this type of survival bias—an important objective when the goal is understanding survival—I focus the analysis on information available for all recruits at the time of enlistment. This includes age, height, occupation, nativity, and place of enlistment.

3.3. Defining demographic groups

Using the data available at the time of enlistment, I first create four samples for analysis. The new urban is key to this analysis. It contains an oversampling of companies that were targeted for sampling because of the high number of urban enlistments found therein. The five targeted cities were New York (including Brooklyn), Philadelphia, Baltimore, Boston, and Chicago. The complete urban collection consisted of over 12,000 soldiers, but not all are from the largest cities. I use here only those from the large cities.

Drawing on the main Union Army sample, the urban oversample, and the USCT sample, I define four mutually exclusive groups. In each case, the sample is constrained to ages 20–39 at enlistment (see discussion on height below), containing complete enlistment data. The groups are:

1 *White rural veterans*. These data include all observations from the main Union Army sample. "Rural" means all veterans not enlisting in a city of more than 10,000 persons. Sample size: 14,931
2 *White urban veterans: Mid-size cities.* These data include observations from the main sample who are from cities that range in size from 10,000 to 100,000 in 1860. (This group does *not* include veterans from the urban oversample). Sample size: 3, 856
3 *White urban veterans: Large cities*. These are soldiers enlisting in the five target cities They come from both the main Union Army sample and the urban oversample. Sample size: 3,798

Table 1
Mean Height (in.) by Enlistment Age.

Age	White, Rural	White, Urban, Mid-size Cities	White, Urban, Large Cities	Black
18	67.04	66.38	65.96	65.55
	2,872	*550*	*465*	*2,261*
19	67.70	67.17	66.22	66.11
	2,220	*510*	*496*	*1,560*
20	67.98	67.29	66.83	66.48
	2,017	*447*	*371*	*1,648*
21	68.09	67.39	66.88	66.60
	1,879	*489*	*497*	*1,578*
22	68.13	67.46	66.91	66.90
	1,587	*374*	*393*	*1,375*
23	68.05	67.30	66.71	66.96
	1,389	*341*	*316*	*1,151*
24	68.23	67.70	66.89	66.89
	1,246	*350*	*282*	*970*
25	68.21	67.57	66.86	67.10
	1,121	*274*	*284*	*971*

Notes: "Rural" means an enlistment city of less than 10,000 in 1860; "Mid-size" cities have 1860 population between 10,000 and 100,000; "Large" cities are New York, Philadelphia, Baltimore, Boston, and Chicago. Height is measured at time of enlistment.

4 *Black veterans.* All these soldiers are from the combined samples collected from companies from the USCT. Sample size: 10,173

Union Army companies were created under the authority of the states and most recruits in a typical company were from the same local area. USCT companies, on the other hand, were formed by the federal government and would draw from many areas, often as the Army made its way through the South liberating slaves. Blacks enlisted in many areas, but since most black soldiers were former slaves from rural areas in the South, I draw no distinction with respect to population for the black analysis sample. Furthermore, the division above does not include all subsets of the collection. In particular, none of the veterans from the urban sample who enlisted outside of the largest cities were used, and those veterans from the main sample who enlisted in cities over 100,000 but were not included as sample target cities, are also not used (the latter cities are New Orleans, Cincinnati, and St. Louis).

All analyses below report results separately by group. This allows us to avoid the imposition of a common height effect across these groups which are likely to be very different—and which results show *are* very different. In theory, combining the large cities that come from two different collections could cause statistical problems, since observations in the urban sample (which is an oversampling) have a greater probability of selection. However, a comparison of the average height between large-city enlistees from the main sample and those from the urban sample shows essentially no difference; thus, they are used together.[4] The same is not true of those from rural areas and mid-size cites. In those cases, the urban sample and the main sample differ in mean height, so the recruits in the urban sample from those areas are not included.

3.4. Height

Height was almost always recorded in the enlistment records of recruits. It is measured in feet and inches, and significant clumping occurs around whole numbers. Because of this, the height variable used below is measured in inches, and dummy variables for specific heights are also measured in inches. Because growth for males can continue into the early twenties, many of the recruits had not completed their growth when they enlisted (and an unknown number lied about their age to be eligible for service, which further understates their growth). Thus, a significant question faces the researcher: what should be the enlistment age cutoff for analysis? Each year the cutoff is increased raises the precision for estimating maximum adult height because more men who have not reached their maximum height are excluded, but each additional year also imposes drastic reductions in sample size, which will reduce precision.

Table 1 illustrates this tradeoff. It shows mean enlistment height at different ages for each of the four demographic groups and the number of observations for each height/age group. In these samples, men are about an inch taller at age 20 than they are at age 18, a big difference. But then the differences taper off quickly (and a portion of the measured increase in height by age occurs because the mid-century decline in heights has just begun).[5] In this analysis, I include recruits who were aged 20 and older in age because the difference in mean height between 20 and 23 within each demographic group is quite small and the gain in sample size from retaining those aged 20–22 is large. I have also done sensitivity tests to see if raising the age cutoff (to 23 or 25) changes the basic results reported below. It does not. I also do not make an adjustment for what is referred to as “shortfall,” the problem caused by minimum height requirements imposed by the Army. This is because estimating the distribution of heights is not my primary objective here, though the possibility exists that not having complete data on the shortest volunteers will introduce some bias into the analysis.

Given these restrictions, Table 2 shows differences in height across the four demographic groups defined above. For the purposes of this analysis, I also differentiate between those who were foreign and native born. Given the large sample sizes, sharp point estimates are achievable, and it is easy to see how mean height varies across demographic sub-groups. About 1.8 in. (4.5 cm) separate the highest group (white, rural, native-born) from the smallest group (white, foreign born in the large cities). Blacks are significantly shorter than their rural white counterparts, but they have almost exactly the same height as white soldiers enlisting in the largest cities, and their height actually exceeds the estimate of the foreign-born whites in the big cities. In short,

[4] Imposing sample weights to account for the different probabilities of selection would be desirable, but since we don’t have an estimate of how many recruits were from the large cities, it is not obvious how to derive appropriate weights.

[5] Komlos (2012) notes how Gould (1869, 104)) measured recruits and thought they grew until age 30. As would be discovered a century later, height was actually declining over time, not with age. In other words, he interpreted a cohort effect as an age effect.

Table 2
Height (in.) at Enlistment (Age 20–39).

	N	Mean	95% C.I.		% Short	% Tall
Rural, White Veterans	17,823	**68.12**	68.08	68.16	8.4%	16.1%
Native	13,428	**68.45**	68.41	68.49	6.2%	18.7%
Foreign-born	4,395	**67.11**	67.04	67.18	15.1%	7.8%
Urban, White Veterans: Mid-size Cities	4,334	**67.51**	67.43	67.58	12.3%	10.6%
Native	2,776	**67.94**	67.85	68.03	8.4%	13.2%
Foreign-born	1,558	**66.74**	66.62	66.86	19.4%	6.0%
Urban, White Veterans: Large Cities	4,413	**66.91**	66.84	66.99	17.5%	6.6%
Native	1,807	**67.28**	67.17	67.76	12.7%	8.1%
Foreign-born	2,606	**66.66**	66.56	66.76	20.8%	5.6%
Black Veterans	13,322	**66.94**	66.90	66.99	18.3%	7.6%

Notes: "Rural" means an enlistment city of less than 10,000 in 1860; "Mid-size" cities have 1860 population between 10,000 and 100,000; "Large" cities are New York, Philadelphia, Baltimore, Boston, and Chicago. Height and nativity are measured at enlistment. "Short" means < 65 in.; "Tall" means ≥ 71 in.

urbanization, nativity and race are all strongly correlated with height.

3.5. Modelling mortality

Waaler's (1984) results suggested that the gradient between height and risk of death falls significantly with age. Thus we would expect to see the largest effects of height in the decades immediately following the war. Unfortunately, this is the time during which death dates are generally not available because of the features of the pension system discussed above. The portion of recruits known to have survived the war that do not have a known death date is 30% for rural whites; 44% for whites in mid-size cities; 58% for whites in large cities; and 48% for blacks. These missing data make any mortality analysis prior to pension liberalization highly tenuous. I, therefore, focus in this paper on veterans who were known to have survived until age 60.

Table 3 summarizes several measures of later-life mortality for each demographic group. As noted before, mortality is positively related to urbanization among white veterans. At age 60, rural white veterans live, on average, another 15.56 years, while those who enlisted in large cities can expect to live another 13.28 years. Blacks live slightly longer than those urban whites. Similar patterns exists for 10-year and 20-year mortality rates at age 60 and for life expectancy and mortality at age 70 among the same group of veterans. These are cohort values and are not directly comparable to values coming from period life tables in the same period.

3.5.1. Explanatory variables

The enlistment records contain the date and place of enlistment, height, occupation, and the state or country of birth. Since status as a farmer has proved an important correlate of height in previous research (Wilson and Pope, 2003), I use a dummy variable for whether or not the recruit was a farmer in his enlistment occupation (non-responses are treated as non-farmers). I also include a dummy variable indicating whether the recruit was native born or not.

Height is the primary variable of interest in this analysis. As noted earlier height is measured at enlistment, and recruits under the age of 20 were eliminated from the analysis because of the significant growth that occurs in the late teens. In general, taller height is thought to reduce mortality, but the approach used here allows the height-mortality relationship to take a variety of shapes and to differ across demographic groups. To accomplish this, I use a set of dummy variables putting heights into ranges. The shortest category are those under 65 in., and the highest category includes those who are 71 in. or taller. The goal is to have the categories in the tails of the distribution be small enough to capture effects at very low or very high height but large enough to have sufficient data points to ensure statistical power.

From other parts of the military file we observe wartime health information on illnesses and wounds and the cause of discharge. In this analysis I use a set of dummy variables to summarize that wartime information: 1) whether the recruit had a record of hospitalization for a wound, 2) whether he had a record of hospitalization for illness; and 3) whether he was discharged for medical reasons. Early discharge may reflect hardship during the war, but it also might mean reduction of health risks associated with service, including further wounds and injuries. Table 4 gives descriptive statistics for the independent variables used in the regression analysis. Mean heights differ from those shown in Table 2 because they include only those cases used in the regression analysis.

One valuable feature of the Union Army data is that companies were drawn from specific localities. In other words, recruits tended to serve with soldiers from their hometowns. Therefore, many of the environmental factors that might affect mortality, such as access to food, local economic conditions, and exposure to disease, would have been shared to some extent by members of the company. Even though we do not observe those variables directly,

Table 3
Later-Life Mortality Measures, by Demographic Group.

	Demographic Group (at Enlistment)											
	White, Rural			White, Urban, Mid-size Cities			White, Urban, Large Cities			Black		
Mortality Measure	N	Mean	Std. Err.	N	Mean	Std. Err.	N	Mean	Std. Err.	N	Mean	Std. Err.
Life Expectancy, Age 60	5,007	**15.562**	0.116	1,034	**14.886**	0.253	636	**13.282**	0.314	2,282	**13.525**	0.166
10-Year Mortality Rate, Age 60	5,007	**0.276**	0.006	1,034	**0.306**	0.014	636	**0.366**	0.019	2,282	**0.365**	0.010
20-year Mortality Rate, Age 60	5,007	**0.692**	0.007	1,034	**0.724**	0.014	636	**0.786**	0.016	2,282	**0.787**	0.009
Life Expectancy, Age 70	3,627	**9.416**	0.102	718	**9.072**	0.220	403	**8.019**	0.284	1,448	**8.157**	0.157
10-year Mortality Rate, Age70	3,627	**0.575**	0.008	718	**0.603**	0.018	403	**0.663**	0.024	1,448	**0.665**	0.012

Notes: Includes only those individuals with known death dates who enlisted between the ages of 20–39 and turned 60 between 1895-1905."Rural" means an enlistment city of less than 10,000 in 1860; "Mid-size" cities have 1860 population between 10,000 and 100,000; "Large" cities are the five largest cities in 1860: New York, Philadelphia, Baltimore, Boston, and Chicago. Height measured at enlistment.

Table 4
Sample Proportions.

	Demographic Group (at Enlistment)							
	White, Rural		White, Urban, Mid-size Cities		White, Urban, Large Cities		Black	
Dependent Variable:	Mean	Std. Err.	Mean	Std. Err.	Mean	Std. Err.	Mean	Std. Err.
10-Year Mortality at Age 60	**0.276**	*0.006*	**0.306**	*0.014*	**0.366**	*0.019*	**0.365**	*0.010*
Independent Variables								
Height (in.) < 65	0.072	*0.004*	0.103	*0.009*	0.156	*0.014*	0.212	*0.009*
65 ≤ Height (in.) < 67	0.189	*0.006*	0.244	*0.013*	0.305	*0.018*	0.279	*0.009*
67 ≤ Height (in.) < 69	0.320	*0.007*	0.313	*0.014*	0.308	*0.018*	0.284	*0.009*
69 ≤ Height (in.) < 71	0.258	*0.006*	0.225	*0.013*	0.173	*0.015*	0.152	*0.008*
71 ≤ Height (in.)	0.162	*0.005*	0.115	*0.010*	0.058	*0.009*	0.073	*0.005*
Native Born	0.846	*0.005*	0.794	*0.013*	0.601	*0.019*		
Farmer	0.629	*0.007*	0.464	*0.016*	0.086	*0.011*		
Any Wartime Illness	0.723	*0.006*	0.625	*0.015*	0.613	*0.019*	0.599	*0.010*
Any Wartime Injury	0.319	*0.007*	0.275	*0.014*	0.379	*0.019*	0.151	*0.007*
Early Disability Discharge	0.194	*0.006*	0.174	*0.012*	0.231	*0.017*	0.072	*0.005*
Sample Size:	5,007		1,034		696		2,282	

Notes: Includes only those individuals with known death dates who enlisted between the ages of 20–39 and turned 60 between 1895-1905."Rural" means an enlistment city of less than 10,000 in 1860; "Mid-size" cities have 1860 population between 10,000 and 100,000; "Large" cities are the five largest cities in 1860: New York, Philadelphia, Baltimore, Boston, and Chicago. Height measured at enlistment.

we can exploit the geographic variation by using company-level fixed effects as a proxy for those important variables.

3.5.2. Linear probability models of 10-year mortality

I employ simple linear probability models to estimate mortality after age 60. Linear probability models are unbiased and their coefficients are easy to interpret. Because the dependent variable to be measured does not have a mean near 0 or 1, little advantage would be gained by using logistic or probit regression. To limit the potential of confusing period and cohort effects, I constrain the sample to those who turned 60 between 1895 and 1904. A 21-year-old recruit at the start of the war would turn 60 in or near 1900. This age range, thus, allows us to capture a fairly wide swath of the sample with enlistment ages between 20–39. All models use heteroscedasticity-robust standard errors, clustered at the company level.

3.5.3. Cox proportional hazard models

I then estimate mortality using a standard Cox proportional hazard model that follows veterans from age 60 until death (only those with known death dates are used). The hazard rate is the probability of dying at a point in time conditional on living up to that point. The model assumes the following form for the hazard at time t for a person in group i:

$$h_i(t) = h_{0i}(t)\, \exp(X_i'B)$$

The Cox model assumes that individuals in group i have a common "baseline" hazard, $h_{0i}(t)$. The matrix of covariates X_i is multiplied by a vector, B, of regression coefficients, which is the linear piece of this specification. A strength of this specification is that no functional form whatsoever is imposed on the baseline hazard rate. The central assumption of the model is that a change in covariates shifts the baseline hazard proportionally. Furthermore, since I estimate the model separately by group, the baseline hazards can follow a different path over time for each group, and the height (and other) coefficients are estimated separately across models. Following convention, I report the results of the regression models in terms of hazard ratios. For instance, a hazard ratio of 1.1 would indicate that a one unit increase in the independent variable results in a 10% increase in the hazard rate..

4. Results

4.1. Linear probability models

Table 5 shows the linear probability estimates of 10-year mortality, estimated separately by demographic group with

Table 5
10-year Mortality: Linear Probability Models

	White, Rural			White, Mid-Size Cities			White, Large Cities			Blacks	
Indep. Variables	Coeff.	Robust S.E.		Coeff.	Robust S.E.		Coeff.	Robust S.E.		Coeff.	Robust S.E.
Height (in.) < 65	0.074	*(.026)*	***	0.044	*(.051)*		−0.039	*(.061)*		−0.015	*(.038)*
65 ≤ Height (in.) < 67	0.026	*(.019)*		−0.008	*(.046)*		0.031	*(.063)*		0.017	*(.034)*
67 ≤ Height (in.) < 69	0.028	*(.015)*	*	0.017	*(.043)*		0.022	*(.055)*		−0.007	*(.035)*
69 ≤ Height (in.) < 71	*Reference*			*Reference*			*Reference*			*Reference*	
71 ≤ Height (in.)	0.058	*(.020)*	***	−0.008	*(.053)*		0.072	*(.104)*		0.042	*(.045)*
Native Born	−0.023	*(.022)*		−0.024	*(.048)*		0.056	*(.049)*			
Farmer	−0.027	*(.015)*	*	0.026	*(.034)*		−0.059	*(.079)*			
Wartime Illness	−0.001	*(.016)*		0.094	*(.041)*	**	0.069	*(.048)*		−0.010	*(.034)*
Wartime Wound	0.041	*(.013)*	***	−0.003	*(.039)*		0.072	*(.048)*		0.086	*(.038)*
Disability Discharge	−0.009	*(.016)*		−0.084	*(.047)*	*	−0.065	*(.051)*		0.335	*(.042)*
Constant	0.260	*(.036)*	***	0.276	*(.091)*	***	0.316	*(.086)*	***	0.335	*(.042)*
N	5,007			1,034			636			2,282	
R-Squared	0.010			0.012			0.011			0.007	
Variance due to fixed effects	0.121			0.305			0.337			0.099	

Notes: Includes only those individuals with known death dates who enlisted between the ages of 20–39 and turned 60 between 1895-1905."Rural" means an enlistment city of less than 10,000 in 1860; "Mid-size" cities have 1860 population between 10,000 and 100,000; "Large" cities are the five largest cities in 1860: New York, Philadelphia, Baltimore, Boston, and Chicago. Height measured at enlistment. Controls for year of enlistment and year turning 60 are included but not presented here (they can be seen in the technical appendix). All models include company-level fixed effects. Standard errors are clustered at the company level. ***: p < .01; **: p < .05; *: p<.1.

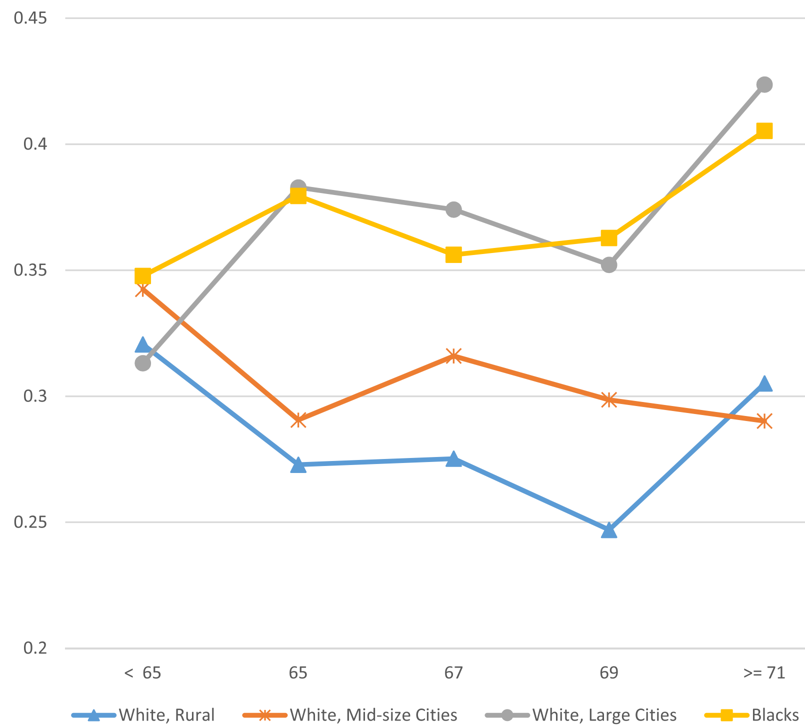

Fig. 2. 10-Year Mortality by Demographic Group and Height (Model Estimates).
Notes: Estimates based on the fixed-effects linear probability models in Table 5. Includes only those individuals with known death dates who enlisted between the ages of 20–39 and turned 60 between 1895-1905."Rural" means an enlistment city of less than 10,000 in 1860; "Mid-size" cities have 1860 population between 10,000 and 100,000; "Large" cities are the five largest cities in 1860: New York, Philadelphia, Baltimore, Boston, and Chicago. Height measured at enlistment.

company-level fixed effects. The first block of estimates reflect the effects of height. For the largest groups, rural whites, we see that both shortness (height <65 in.) and tallness (height ≥71 in.) have higher mortality, and those estimates are highly statistically significant. The general pattern across the height distribution for this group is that mortality falls with greater height—except for the strong uptick at the upper end of the distribution. This pattern, however, does not hold for the other demographic groups. At this point it cannot be known whether this is because the sample sizes are not large enough or whether the true effects of height are very small for this stage of the life-cycle.

Fig. 2 shows predicted mortality probabilities for each of the four groups based on the models shown in Table 5, holding all values of independent variables fixed at mean levels. The estimated pattern for rural whites is fairly similar to Costa (1993), who also uses white veterans, and Waaler's (1984) estimates on the Norwegian population. The mean differences in 10-year mortality between groups shown earlier in Fig. 1 (in other words, the gaps between the trend lines in Fig. 2) are statistically significant, but the slopes of the trend line for all groups except rural whites are not significant. Still, a striking feature of this figure is that the higher mortality populations—whites from large cities and blacks—are very similar to each other but very different from the lower mortality groups.

Substantively, the height profiles shown in Fig. 2 are non-trivial. In particular, the mortality estimate for the shortest group of rural whites (.321) is 30% higher than the mortality (.247) for those with height between 69 and 71 in., after controlling for the effect of other variables. By itself, those estimates lead to the conclusion that height is related to later life mortality, at least in some cases. Furthermore, Fig. 2 confirms the importance of not imposing a linear relationship (or other functional form) for height and mortality. In exploratory analysis represented in the technical appendix, I use a simple linear model in height and find that for rural whites, the coefficient on height is -.001 ($p = .681$), which is both small and statistically insignificant. The same non-effects are found in other groups.

The other independent variables don't reveal any general patterns in the variables that we think would affect later life mortality. Among the rural white population, native born recruits and farmers have lower mortality, but not significantly so. Having a war wound raises mortality by over 4 percentage points, but illness and disability discharge have no effect. For the other groups, estimates are generally imprecise and insignificant, and the direction of the effects is too haphazard to detect any sort of pattern. The same holds for year of enlistment and for year turning 60.

Finally, in theory the fixed effects in the model account for the location-specific unobserved factors affecting mortality, such as local disease environment or food quality. And Table 5 shows that the fixed effects do account for some of the explained variation in mortality.[6] However, inclusion of the fixed effects has almost no impact on the coefficient estimates for height. Estimates for models without the fixed effects are presented in the technical appendix.

4.2. Cox model estimates

The final step in the analysis is to estimate mortality through the end of the lifespan using Cox proportional hazard models. Table 6 reports estimates from these models, with coefficients represented by hazard ratios. For the white, rural veterans, the hazard of mortality is slightly higher for shorter heights, but the estimate of 1.06 is not statistically significant. The increasing hazard for greater height is more apparent, with the top value at 71 in. or higher reaching a significant estimate of 1.12. In other

[6] 12% for rural whites; 31% of the mid-size city group, 34% for the large city group, and 10% for the African-American group.

Table 6
Cox Proportional Hazard Models.

	White, Rural			White, Mid-Size Cities		White, Large Cities			Blacks	
Indep. Variables	Haz. Ratio	Robust S.E.		Haz. Ratio	Robust S.E.	Haz. Ratio	Robust S.E.		Haz. Ratio	Robust S.E.
Height (in.) < 65	1.052	*(.068)*		0.929	*(.102)*	0.838	*(.098)*		1.037	*(.074)*
65 ≤ Height (in.) < 67	0.993	*(.040)*		0.998	*(.080)*	0.964	*(.124)*		1.042	*(.064)*
67 ≤ Height (in.) < 69	1.013	*(.034)*		1.059	*(.075)*	1.067	*(.127)*		1.019	*(.069)*
69 ≤ Height (in.) < 71	*Reference*			*Reference*		*Reference*			*Reference*	
71 ≤ Height (in.)	1.128	*(.048)*	***	1.083	*(.102)*	1.011	*(.173)*		1.187	*(.106)*
Native Born	0.983	*(.041)*		0.900	*(.059)*	1.036	*(.088)*			
Farmer	0.845	*(.026)*	***	0.897	*(.060)*	0.831	*(.128)*			
Wartime Illness	0.930	*(.031)*	**	1.194	*(.077)*	1.200	*(.106)*	**	0.965	*(.044)*
Wartime Wound	1.047	*(.031)*		0.943	*(.068)*	0.984	*(.084)*		1.003	*(.062)*
Disability Discharge	0.975	*(.034)*		0.902	*(.082)*	0.981	*(.090)*		1.173	*(.091)*
N	5,007			1,034		636			2,282	

Notes: Includes only those individuals with known death dates who enlisted between the ages of 20–39 and turned 60 between 1895-1905."Rural" means an enlistment city of less than 10,000 in 1860; "Mid-size" cities have 1860 population between 10,000 and 100,000; "Large" cities are the five largest cities in 1860: New York, Philadelphia, Baltimore, Boston, and Chicago. Height measured at enlistment. Controls for year of enlistment and year turning 60 are included but not presented here (they can be seen in the technical appendix). Standard errors are clustered at the company level. ***: p < .01; **: p < .05; *: p<.1.

words the basic U-shape of the height-mortality relationship is still present in the Cox model, but it is only the taller recruits who have significantly greater mortality (however, the U-shaped pattern is more pronounced and statistically significant in the results presented below in Section 4.3). The other control variables in the model such as nativity and occupation have no significant effects.

The height-mortality pattern seen among rural whites is not found for the other demographic groups in Table 6. In fact, to the extent that any pattern exists in these results, it is one of a positive relationship between height and mortality. This has been found before. Su (2009) estimated a proportional hazard model with a linear, continuous specification for height. He finds a very small (HR = 1.01) but significant and *positive* effect of increasing height on mortality. Costa (2015) also mentions her research (not presented in the paper) that shows a "J-shaped" height gradient, in other words, mortality increases at taller heights. And Waaler's (1984) estimates showed an uptick in mortality at the upper end of the height distribution, even though the relationship was negative at lower heights.

4.3. Sensitivity analysis

The decision to include veterans in this analysis whose enlistment age (and, hence, the date of their height measurement) was 20 or older is potentially problematic. This is because men sometimes have significant growth beyond their early 20 s in historical populations (Beekink and Kok, 2017). The cost of increasing the minimum age, however, is a large reduction in sample sizes and the associated loss of precision and statistical power. Such a restriction is compounded by the restriction that the mortality analysis includes only those cases who turn 60 after Jan. 1, 1895. I imposed that restriction because of the need to ensure that the baseline group of veterans turn 60 well after the liberalization of the law in 1890.

Table 7 shows estimates of the height effects from a regression that includes only those men aged 24–39 at recruitment and compares them to the earlier estimates. Despite the loss in sample size, the estimated effects of height for rural whites are much more pronounced for both the 10-year mortality estimates and in the hazard models. For example, 10-year mortality is almost 15 percentage points (.146) higher among the shortest height category than it is in the reference category. This is a very large effect given that the average mortality among rural whites is 27.6% (see Table 4). Similarly the mortality for the tallest group is 8.5 percentage points higher than the reference, indicating that higher 10-year mortality is found among both the shortest and tallest categories among rural whites. Interestingly, the coefficient estimates for whites from smaller cities are very similar in magnitude (.153 for the shortest and .099 for the tallest) to the rural whites, but those estimates are not statistically significant.

Table 7 also shows that the Cox model estimates for rural white are much different than the weak results found earlier in Table 6. Consistent with the U-shaped pattern for the 10-year mortality results just discussed, the hazard of mortality is significantly higher for the shortest and tallest categories (HR = 1.298 for the shortest category and HR = 1.151 for the tallest). For the other demographic groups, a statistically significant pattern for height is not found in the Cox model estimates.

In the sensitivity analysis shown in Table 7, in which the sample of recruits is restricted to those age 24 or older at enlistment, mortality at both the shortest and tallest heights is higher than the estimates in Tables 5 and 6. This suggests attenuation bias in the results due to measurement error induced by obtaining heights before growth is complete. In these latter results, whites from mid-size cities have a similar (but insignificant) height-mortality relationship to rural whites.

5. Discussion

Komlos states, "From a theoretical perspective, we know that height is a positive function of income and in every single data set examined we do find that wealthier parents have taller children, everything else being equal. There is absolutely no exception to this generalization as long as there are no simultaneous offsetting effects" (2012, 396). A similarly strong relationship exists between income and mortality, though much remains to be understood about why that gradient exists and why it differs across populations (Chatty et al., 2016). Together, these correlations suggest a negative relationship between height and mortality.

However, the analysis here does not indicate a robust negative relationship across all height categories—especially when comparing results across the demographic groups. I do find that, for rural whites, the shortest veterans have the highest mortality. For example, the 10-year mortality model in Table 7 that restricts the sample to those aged 24 and older (a more conservative approach), predicts a mortality rate of .374 for those under 65 in. compared to only .228 for those in the 69–71 inch group, indicating that mortality is 64% higher among the shortest group than it is among the reference group, a very large difference. But mortality is also

Table 7
Sensitivity Analysis for Age at Enlistment.

	10-Year Mortality (LPM)						Cox Proportional Hazard Model				
	Full Model:			*Restricted Model:*			*Full Model:*		*Restricted Model:*		
	Enlist. Age: 20-39			*Enlist. Age: 24-39*			*Enlist. Age: 20-39*		*Enlist. Age: 24-39*		
White, **Rural**											
Height Category	Coeff.	Robust S.E.		Coeff.	Robust S.E.		Haz. Ratio	Robust S.E.	Haz. Ratio	Robust S.E.	
Height (in.) < 65	0.074	*(.026)*	***	0.146	*(.050)*	***	1.052	*(.065)*	1.298	*(.128)*	***
65 ≤ Height (in.) < 67	0.026	*(.019)*	*	0.055	*(.031)*	*	0.993	*(.040)*	1.059	*(.071)*	
67 ≤ Height (in.) < 69	0.028	*(.015)*	*	0.024	*(.026)*		1.013	*(.034)*	1.021	*(.051)*	
69 ≤ Height (in.) < 71											
71 ≤ Height (in.)	0.058	*(.020)*	***	0.085	*(.032)*	***	1.128	*(.048)*	1.151	*(.080)*	**
N=	5,007			2,036			5,007		2,036		
White, **Mid-Size Cities**											
Height Category	Coeff.	Robust S.E.		Coeff.	Robust S.E.		Haz. Ratio	S.E.	Haz. Ratio	Robust S.E.	
Height (in.) < 65	0.044	*(.051)*		0.153	*(.083)*	*	0.929	*(.102)*	1.118	*(.187)*	
65 ≤ Height (in.) < 67	−0.008	*(.046)*		0.068	*(.070)*		0.998	*(.080)*	1.108	*(.139)*	
67 ≤ Height (in.) < 69	0.017	*(.043)*		0.062	*(.061)*		1.059	*(.075)*	1.177	*(.138)*	
69 ≤ Height (in.) < 71											
71 ≤ Height (in.)	−0.008	*(.053)*		0.099	*(.075)*		1.083	*(.102)*	1.289	*(.187)*	*
N=	1,034			474			1,034		474		
White, **Large Cities**											
Height Category	Coeff.	Robust S.E.		Coeff.	Robust S.E.		Haz. Ratio	Robust S.E.	Haz. Ratio	Robust S.E.	
Height (in.) < 65	−0.039	*(.061)*		0.065	*(.143)*		0.838	*(.098)*	0.942	*(.224)*	
65 ≤ Height (in.) < 67	0.031	*(.063)*		0.141	*(.122)*		0.964	*(.124)*	1.133	*(.230)*	
67 ≤ Height (in.) < 69	0.022	*(.055)*		0.142	*(.105)*		1.067	*(.127)*	1.266	*(.243)*	
69 ≤ Height (in.) < 71											
71 ≤ Height (in.)	0.072	*(.104)*		−0.024	*(.148)*		1.011	*(.173)*	1.454	*(.478)*	
N=	636			245			636		245		
Blacks											
Height Category	Coeff.	Robust S.E.		Coeff.	Robust S.E.		Haz. Ratio	Robust S.E.	Haz. Ratio	Robust S.E.	
Height (in.) < 65	−0.015	*(.038)*		−0.011	*(.056)*		1.037	*(.074)*	1.031	*(.102)*	
65 ≤ Height (in.) < 67	0.017	*(.034)*		0.009	*(.048)*		1.042	*(.064)*	1.061	*(.093)*	
67 ≤ Height (in.) < 69	−0.006	*(.035)*		0.010	*(.047)*		1.019	*(.069)*	1.130	*(.097)*	
69 ≤ Height (in.) < 71											
71 ≤ Height (in.)	0.043	*(.045)*		0.028	*(.066)*		1.187	*(.106)*	1.150	*(.154)*	
N=	2,282			1,053			2,282		1,053		

Notes: Includes only those individuals with known death dates who turned 60 between 1895-1905."Rural" means an enlistment city of less than 10,000 in 1860; "Mid-size" cities have 1860 population between 10,000 and 100,000; "Large" cities are the five largest cities in 1860: New York, Philadelphia, Baltimore, Boston, and Chicago. Height measured at enlistment. Controls for year of enlistment and year turning 60 are included but not presented here (they can be seen in the technical appendix). Standard errors are clustered at the company level. ***: p < .01; **: p < .05; *: p<.1.

much (37.3%) higher among the tallest group (.313 compared to .228) than it is among the reference category. Cox models, which follow the veterans from age 60 to death, show a similar pattern, though somewhat less pronounced.

What could explain this U-shaped nature of mortality among rural whites? One possible explanation might have to do with *when* the nutritional deficits occur. Shorter individuals include a subset of people who are genetically predisposed to be taller but are stunted because of nutritional deficits during growth periods in childhood and adolescence. Tall people, on the other hand, have a higher nutritional demand throughout their lives, holding other factors constant. Thus, in an environment where food is scarce, taller people will face a greater nutritional deficit. If this deficit persists over time, it may lead to biological consequences that affect mortality in later life. Thus, according to this conjecture, short people would have higher mortality because, as a group, they were disproportionally stunted during childhood and adolescence. Tall people, on the other hand, would be disproportionally undernourished as adults, perhaps for many years.

Previous historical research has paid scant attention to the potential effects of being taller than average, though the effect is seen in Costa (1993), Waaler (1984) and Su (2010). In additional to the nutritional conjecture, perhaps social factors related to being tall had an effect on mortality; or perhaps diseases more common among the tall (Batty et al., 2009) were important. Research to date has little to say about these possibilities. The international evidence summarized in Section 2.2 also found a mix of positive and negative effects with respect to height and a host of null effects—not unlike the results presented here

Proving this conjecture would require data that matched nutritional resources, in a highly specific manner, to both time and place. Such an analysis would go well beyond the scope of this paper and may not be feasible with any available dataset. However, the analysis in this paper does go part way in accounting for local, time-specific factors through the use of company-level fixed effects in the regressions above. Recruits in the same company shared, to a significant extent, a common childhood environment, and they were approximately the same age. As shown above, the company-level effects explain a small but non-trivial part of later life mortality.

A second lingering question is why are significant height effects found only among rural whites and not the other demographic groups. A major innovation of this analysis is to compare individual-level effects of height across demographic groups defined by race and level of urbanization. My intuition was that blacks and urban whites would have a gradient negative and steeper than that found for rural whites, since those subpopulations faced both a lower nutritional resources and a higher disease burden. But the analysis did not support this intuition, since the height effects are found only among the rural whites.

Perhaps larger sample sizes would confirm the findings merely suggested by Fig. 2, that whites from large cities and blacks have a generally increasing gradient. A similar nutritional conjecture as that discussed above above could explain the positive gradient. In this case, long-term effects of nutritional deficiencies caused by the higher nutritional demands of taller individuals could make the taller individuals subject to higher mortality in late adulthood. Gaining local ecological data for black soldiers is daunting, however, and USCT companies were not drawn from local areas to the same extent that white companies were.

The analysis here focuses on information found in the military service records, including enlistment data and the health outcomes during the war because this set of variables is nearly comprehensive and not subject to pension-related selection bias and because census linkages in the *Early Indicators* data are highly dependent on survival to later ages. Clearly, there may be other intervening variables through which height could affect later mortality. Factors such as income, occupation, residential history and marital status are candidates for such intervening factors. For rural whites, the additional of variables from the census and pension files may help account for the pattern of height effects found here. For other demographic groups, there is no association between height and mortality that one could seek to explain through the addition of these intervening variables, thus searching for intervening variables is unlikely to be a fruitful exercise.

6. Conclusions

A central finding of anthropometric research is that mean height in a population reflects nutritional history and, hence, the health of a population. This is the essential story told by Fig. 1 at the outset of this paper. In that figure, height varies strongly and significantly with respect to race, urbanization, and nativity, all factors that predict mortality. Also, much previous research has shown that over long periods of time in the U.S., height has tended to move together with life expectancy (Costa, 2015).

But how far can we stretch the relevance of height as a measure of population health? In particular, is adult height a significant predictor of mortality at the individual level? In this paper, I have used large and newly collected data on blacks and urban white veterans of the Civil War to supplement the existing data on white veterans. I have then searched for a height gradient within demographic groups defined by race and urbanization. The main empirical finding here is that among rural whites, a statistically significant U-shaped pattern exists, with higher 10-year mortality for the shortest (height <65 in.) and tallest (height $\geq$ 71.in) veterans (See Fig. 2 for reference).

The pattern mentioned above is not, however, found among whites from large cities nor among blacks (even though these two groups have significantly lower height and life expectancy than the other groups (see Fig. 1 and Tables 2 and 3). And if we take at face value the estimates shown in Fig. 2, the other demographic groups are quite distinct from the rural white population. Indeed, if any pattern exists among these groups it is that mortality increases positively with height, which is found (*without* statistical significance) in both the linear probability models and the Cox proportional hazard models.

Trying to predict old age mortality from variables in early life is a challenging enterprise under any conditions. Even though the samples may start out very large, once we impose all the sample restrictions necessary to ward off the selection bias problems in the *Early Indicators* data, the analytical samples are much more modest in size. This is particularly true of the urban oversample. A major reason for this is that so many of these recruits died before age 60. Perhaps different methods could be employed that used larger data sets and achieve, therefore, more precise estimates. However, the African-American sample is about 3 times the size of the large city sample, and the precision of the height effects there are even weaker than in the smaller urban groups.

By modern standards, men who are 71 in. or taller are not exceptionally tall, but those under 65 in. *are* exceptionally short. This raises the question of how we translate the height distribution of these historical populations into modern contexts where the height distributions are so much different. As discussed in Alter (2004), we would not necessarily expect the height-mortality relationship to be stable, given the difference in environmental conditions across historical periods and populations. And it is not clear that shorter populations that exist today in low-income countries will follow the same pattern as found in historical populations that have similar height distributions.

That height varies so significantly across demographic groups in the 19th Century U.S. indicates the importance of variables such as race, urbanization, and nativity in explaining economic growth and its consequences, as illustrated in Fig. 1. The research here uses newly collected data that allows the effect of height to be explored *within* those groups. In sum, even though mean height is a powerful predictor of the health of a population and can explain mortality differences across groups in the population, how height affects mortality at the individual level in historical US populations is still a largely open question. This is the classic ecological problem that group-level relationships may be quite different than individual-level ones.

References

Alter, G., 2004. Height, frailty, and the standard of living: modelling the effects of diet and disease on declining mortality and increasing height. Popul. Stud. 58, 265–279.

Batty, G.D., et al., 2009. Height, wealth, and health: an overview with new data from three longitudinal studies. Econ. Hum. Biol. 7, 137–152.

Batty, G.D., Leon, D.A., 2002. Socio-economic position and coronary heart disease risk factors in children and young people. Evidence from UK epidemiological studies. Eur. J. Public Health 12, 263–272.

Beekink, E., Kok, J., 2017. Temporary and lasting effects of childhood deprivation on male stature. Late adolescent stature and catch-up growth in Woerden (the Netherlands) in the first half of the nineteenth century. Hist. Fam. 22, 196–213.

Chetty, R., Stepner, M., Abraham, S., Lin, S., Scuderi, B., Turner, N., Bergeron, A., Cutler, D., 2016. The Association Between Income and Life Expectancy in the United States, 2001-2014. JAMA 315, 1750–1766.

Costa, D.L., 1993. Height, weight, wartime stress, and older age mortality: evidence from the union army records. Explor. Econ. Hist. 20, 424–449.

Costa, D.L., 2015. Health and the economy in the United States from 1750 to the present. J. Econ. Lit. 53, 503–570.

Costa, D.L., DeSomer, H., Hanss, E., Roudiez, C., Wilson, S.E., Yetter, N., 2017. Union Army veterans, all grown up. Hist. Methods 50, 79–95.

Davey Smith, G., Hart, C., Hole, D., MacKinnon, P., Gillis, C., Watt, G., Blane, D., Hawthorne, V., 1998b. Education and occupational social class: which is the more important indicator of mortality risk? J. Epidemiol. Commun. Health 52, 153–160.

Davey Smith, G., Hart, C., Upton, M., 2000. Height and risk of death among men and women: aetiological implications of associations with cardiorespiratory disease and cancer mortality. J. Epidemiol. Commun. Health 54, 97–103.

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

Fogel, R.W., 1986. Nutrition and the Decline in Mortality Since 1700: Some Preliminary Findings. In Long-term Factors in American Economic Growth. University of Chicago Press, Chicago.

Fogel, R.W., Engerman, S.L., 1974. Time on the Cross: the Economics of American Negro Slavery. Little, Brown., Boston.

Fogel, R.W., Engerman, S.L., Trussell, J., Floud, R., Pope, C.L., Wimmer, L., 1978. The economics of mortality in North America, 1650–1910: a description of a research project. Hist. Methods 11, 75–108.

Gould, B., 1869. Investigations in the Military and Anthropological Statistics of American Soldiers, 2019 Soldiers. Riverside Press, Cambridge, MA.

Grasgruber, P., Sebera, M., Hrazdira, E., Cacek, J., Kalina, T., 2016. Major correlates of male height: a study of 105 countries. Econ. Hum. Biol. 21, 172–195.

Ihira, H., Sawada, N., Iwaski, M., Yamaji, T., Goto, A., Noda, M., et al., 2018. Adult height and all-cause and cause-specific mortality in the Japan Public Health Center-based Prospective Study (JPHC). PLoS One 13 (5) e0197164.

Jousilahti, P., Tuomilehto, J., Vartiainen, E., Eriksson, J., Puska, P., 2000. Relationship of adult height to cause-specific and total mortality: a prospective follow-up study of 31,199 middle-aged men and women in Finland. Am. J. Epidemiol. 151, 1112–1120.

Komlos, J., 2012. A three-decade history of the antebellum puzzle: explaining the shrinking of the U.S. Population at the onset of modern economic growth. J. Hist. Soc. 12, 395–445.

Komlos, J., Kriwy, P., 2002. Social status and adult heights in the two Germanies. Ann. Hum. Biol. 29, 641–648.

Komlos, J., Lauderdale, B.E., 2007. Underperformance in affluence: the remarkable relative decline in U.S. Heights in the second half of the 20th century. Soc. Sci. Q. 88, 283–305.

Rohrmaann, S., Hale, S.R., Staug, K., Bopp, M., Faeh, D., 2017. Body height and mortality – mortality follow-up of four Swiss surveys. Prevent. Med. 101, 67–71.
Sawada, N., Wark, P.A., Merritt, M.A., Tsugane, S., Ward, H.A., Rinaldi, S., et al., 2017. The association between Adult attained height and sitting height with mortality in the European Prospective Investigation into Cancer and Nutrition (EPIC). PLoS One 12 (3) e0173117.
Silventoinen, K., Lahelma, E., Lundberg, O., Rahkonen, O., 2001. Body height, birth cohort and social background in Finland and Sweden. Eur. J. Public Health 11, 124–129.
Steckel, R.H., 1995a. Stature and the standard of living. J. Econ. Lit. 33, 1903–1940.
Sohn, K., 2016. Now, the taller die earlier: the curse of cancer. J. Gerontol. A Biol. Sci. Med. Sci. 71, 713–719.
Song, Y.M., Smith, D., Davey Smith, G., Sung, J., 2003. Adult height and cause-specific mortality: a large prospective study of South Korean men. Am. J. Epidemiol. 158, 479–485.
Steckel, R.H., 1995b. Stature and the standard of living. J. Econ. Lit. 33, 1903–1940.
Su, D., 2009. Risk exposure in Early life and mortality at older ages: evidence from Union Army veterans. Popul. Dev. Rev. 35 (2), 275–295.
Waaler, H.T., 1984. Height, weight and mortality. The Norwegian experience. Acta Medica Scandivica (Supplement) 679, 1–56.
Wilson, S.E., 2003. The prevalence of chronic respiratory disease in the industrial era: the United States, 1895–1910. In: Costa, D.L. (Ed.), Health and Labor Force Participation Over the Life Cycle: Evidence from the Past, 2019 from the Past. NBER and University of Chicago Press, Chicago, pp. 147–180.
Wilson, S.E., 2010. Prejudice and policy: racial discrimination in the Union Army disability pension system, 1865–1906. Am. J. Public Health 100, S56–S65.
Wilson, S.E., Pope, C.L., 2003. The height of Union army recruits: family and community influences. In: Costa, D.L. (Ed.), Health and Labor Force Participation over the Life Cycle: Evidence from the Past, 2019 from the Past. NBER and University of Chicago Press, Chicago, pp. 147–180.