HISTORICAL METHODS
2017, VOL. 50, NO. 2, 79–95
http://dx.doi.org/10.1080/01615440.2016.1250022

Routledge
Taylor & Francis Group

Union Army veterans, all grown up

Dora L. Costa[a,b], Heather DeSomer[b], Eric Hanss[b], Christopher Roudiez[b], Sven E. Wilson[c], and Noelle Yetter[b]

[a]Department of Economics, University of California, Los Angeles; [b]National Bureau of Economic Research; [c]Department of Political Science, Brigham Young University

ABSTRACT

This article overviews the research opportunities made possible by a National Institute on Aging-funded program project, *Early Indicators, Intergenerational Processes, and Aging*. Data collection began almost three decades ago on 40,000 soldiers from the Union Army in the U.S. Civil War. The sample contains extensive demographic, economic, and medical data from childhood to death. In recent years, a large sample of African-American soldiers and an oversampling of soldiers from major U.S. cities have been added. Hundreds of historical maps containing public health data have been geocoded to place soldiers and their family members in a geospatial context. With newly granted funding, thousands of veterans will be linked to the demographic information available from the census and vital records of their children.

KEYWORDS

Health; historical demography; mortality; socioeconomic status

Overview

A growing body of literature in the social sciences and in epidemiology and medicine focuses on the life cycle and intergenerational determinants of death, disease, and socioeconomic status (Costa 2015; Currie 2009; Gluckman et al. 2005). Unfortunately, few datasets permit such studies. Social scientists can study recent cohorts through the HRS and the PSID, but if we wish to understand long-run trends we must examine historical cohorts. Historical datasets, however, tend to be geographically constrained, narrow in scope, and tied to a particular period of time. In contrast, the National Institute on Aging-funded *Early Indicators* project, which traces the life histories of the first U.S. cohort to reach age 65 in the twentieth century, captures demographic, socioeconomic, and health data on a broadly representative sample of tens of thousands of men and their families, linked over time, for a century. Still expanding, the data collections include high-quality individual, household, and company-level data coming from a period of immense economic (industrialization, urbanization) and demographic (mortality and fertility transitions) change.

Data in the *Early Indicators* collections are centered on random samples of over 70,000 soldiers (Black and White, urban and rural, immigrant and native-born) serving in the Union Army (UA) during the U.S. Civil War. The UA veterans are observed in all the surviving federal censuses from 1850–1930s, as children residing with their parents, as soldiers, and later as beneficiaries of a federal pension system that collected detailed health, demographic, and economic data until they died. Supplemental collections have augmented the original sample of almost 40,000 white recruits with over 21,000 African-American soldiers and an oversampling of 12,500 recruits from major cities in nineteenth-century America (in addition to the urban recruits already in the main sample). Because the rate of participation in the UA was so high, the sample is largely representative of the men in the North who were late teens and early twenties during the Civil War. Furthermore, because all data were collected from public sources, there are no confidentiality requirements for using these data, and all data can be accessed through the website *UAdata.org*.

In recent years, the UA sample has been complemented with a vast ecological dataset that contains a century's worth of public health, economic, and demographic detail from large U.S. cities. Researchers have merged thousands of city maps, including detailed street-level changes over time, with thousands of tables of data that are linked to specific urban locations or to wards and neighborhoods. Because urban veterans in the pension system have street addresses in their pension files, veterans can be precisely located and geocoded. This allows analysis not only of urban living patterns but also of the environmental risk factors faced by urban veterans.

CONTACT Sven E. Wilson sven_wilson@byu.edu Department of Political Science, Brigham Young University, 722 SWKT, Provo, UT 84602-0002, USA.
Color versions of one or more figures in this article are available online at www.tandfonline.com/vhim.

Over the next five years, the research team will use powerful new genealogical tools to link tens of thousands of children of UA veterans to their death records and to the surviving manuscript schedules of the federal census from 1860–1940, as well as construct a similar non-UA intergenerational sample from published family histories. These new efforts allow a truly intergenerational and geospatial analysis of demographic processes.

This article illustrates the wide-ranging research potential now available in the *Early Indicators* collection and highlights the exciting new developments and ongoing collections that will soon be available to researchers. Though many papers have been published using the *Early Indicators* collections, in many respects it is a collection that is only now coming of age. In the following sections, we hope to provide an impetus and guide for new research with the *Early Indicators* collections. We outline briefly the history of the project and provide a detailed review of both data that have already come online and those that are in process. We then provide numerous examples of research topics that interested scholars might pursue using the *Early Indicators* data, and we discuss several issues that researchers should be aware of when using the data. Though the project has been underway for three decades, new research opportunities with the *Early Indicators* collections are abundant.

Background

In the 1970s, an international group of scholars began using new data sources to understand long-term mortality trends, including the secular decline in mortality since the 1600s (Fogel et al. 1978). Research by Thomas McKeown (1976, 1978) in the preceding decades had emphasized the role of improved nutrition on mortality decline, which challenged the prior consensus (United Nations 1973) that mortality decline was due to public health, medical knowledge, hygiene, and income. Efforts by scholars working with these new data sources came up with refined evidence on trends and more nuanced explanations of those trends (Fogel 1986; Haines 1979; Pope 1992).

Robert Fogel was a driving force behind much of the new data collection and analysis. He and Stanley Engerman had used data on heights from probate records to study the mortality of slaves in their path-breaking work, *Time on the Cross* (1974). Their work expanded the possibility of using anthropometric measures, such as height and BMI, to study the role of chronic malnutrition on health and longevity, which seemed to be important in explaining the new finding that health and mortality could go through periods of decline even during times of strong economic growth (Fogel 1986; Komlos 1987; Margo and Steckel 1983). As part of the effort to explain that paradox and to further understand secular mortality decline, Fogel envisioned a vast new collection of data based on the military and pension records from the U.S. Civil War found in the National Archives (Wimmer 2003).

The scope of Fogel's proposal was daunting. It involved collecting and digitizing very large archives and linking individuals across data sources. Today, these tasks remain a significant technical and logistical challenge. In the 1980s, it was almost inconceivable since nothing of this scale had ever been done before. A grant application in 1986 was rejected on the grounds that such a massive and wide-ranging data collection project was not feasible. Fogel, ever the optimist, was determined to prove the critics wrong. With assistance from the National Bureau of Economic Research, he began collecting a pilot sample of 20 UA companies, developing at the same time the detailed software and collection methods necessary to capture the soldiers' experiences during the war, their family backgrounds from the federal censuses, and their health histories found in the pension records housed in the National Archives. Armed with the evidence of a successful pilot, Fogel's 1991 application to the National Institute on Aging and the National Science Foundation, *Early Indicators of Later Work Levels, Disease and Death*, was successful, which initiated the large-scale collection of veteran data (Fogel 1993). From the beginning, the project involved scholars from many universities. Larry Wimmer at Brigham Young University oversaw the logistics and training of the data collectors at the National Archives, as well as inputters in Chicago and Provo, Utah. Medical oversight was provided by doctors Nevin Scrimshaw, Irv Rosenberg, and Louis Nguyen (Wimmer 2003).

The *Early Indicators* collections have reshaped knowledge of health and aging. Based on what they were seeing in the veterans' pension records, Robert Fogel and Dora Costa (1997) proposed a new theory of human physiology, which they entitled "technophysio evolution." The theory was that technologically induced changes in the environment had induced a form of human evolution that was "biological but not genetic, culturally transmitted, and not necessarily stable" (49). As the vast medical data in the pension files (referred to as the "surgeon's certificates") started accumulating, it became apparent that these veterans had a host of negative health conditions. At the time, the idea was still in vogue that the elderly in pre-transition societies would be in relatively good health and free from disability because, it was thought, the robust survived and the frail died (Alter and Riley 1989; Myers, Lamb, and Agree 2003; Riley and Alter 1989; Verbrugge 1984). This idea was sharply challenged by the new data showing high rates (relative to

the mid-twentieth century) of disease (Costa 2000) and disability (Costa 2002). The composition of disabling conditions had also shifted considerably given that many non-fatal conditions became less disabling over time because of both improving medical treatment and the declines in the physical requirements of work (Song and Nguyen 2003; Wilson, Burton and Howel 2005; Wilson and Nguyen 1998).

Since the 1990s, scholars have pursued a variety of other important research topics made possible by the UA data. These include the benefits of social networks (Costa and Kahn 2007a, 2007b, 2008, 2010); labor force participation and retirement (Costa 1998; Lee 1998, 2001, 2007); the relationship between wealth accumulation, economic mobility, and health (Lee 2005, 2008); marriage patterns (Hacker 2008); survival in urban environments (Cain and Hong 2009); extreme aging (Costa and Lahey 2005); and the epidemiology, social ecology, and economic consequences of important diseases, such as arteriosclerosis (Costa, Helmchen and Wilson 2007), arthritis (Canavese and Fogel 2009), malaria (Hong 2011), and chronic respiratory disease (Wilson 2003).

UA veterans and related collections

Since the initial successful grant in 1991, the *Early Indicators* project has expanded its scope considerably. In addition to the original sample of over 39,000 white enlisted men from the UA, project researchers have added a sample of 21,000 African-American veterans from the United States Colored Troops (USCT); an oversample of 12,500 urban recruits from the largest U.S. cities; 1,000 POWs from the infamous Andersonville camp; and a collection of 1,700 veterans who lived to at least the age of 95. The project has also added a detailed set of records called the Historical Urban Ecological (HUE) data. These data are collected from thousands of historical maps, which will allow the ability to place veterans in a rich geographical context. Finally, our researchers are now collecting data on the children of UA veterans, both black and white, facilitating an unprecedented opportunity to study the intergenerational aspects of aging. As a complement to that new intergenerational sample, funding will also be used to release publicly the sample of U.S. family history data originally collected by economic historian Clayne Pope. The following sections provide more detail on each of these collections.

Completed veteran collections

UA veterans (UAV)

The main thrust of that first major grant in 1991 was a sample of 39,340 recruits consisting of the complete enlistment of 331 randomly drawn companies in the UA. Recruits in the sample were enlisted white soldiers in infantry units from all Northern states. Wartime data are drawn first from the Compiled Military Service Records (CMSRs), which contain information on the recruit from his enlistment record (residence, occupation, age, height) and variables from a monthly muster record, which includes health, battles, wounds, hospitalization, desertion, POW status, and cause of death or muster-out. These data are supplemented by variables from the Carded Medical Records (CMRs), which contain information from military hospitals. The cluster-sample approach allows researchers to calculate company-level variables, such as the company mortality or ethnic diversity of the company.

From the starting point of the CMSR and CMR records, researchers then go both forward and backward in the soldier's life. Using the index of pensions, they locate the pension records for those soldiers who lived long enough to enter the pension system. From the pension files and the accompanying "Surgeons' Certificates," researchers draw a wealth of information, often covering several decades, which includes demographic and economic information, and the medical records from examinations that were conducted by the Pension Bureau to determine eligibility for pension support. Pension records can be dozens and sometimes hundreds of pages in length, and they constitute the primary source of data in the UA collection.

The final source of records comes from the federal census. When the project began, the most recent publicly-available census was 1910, so the veterans were linked forward to the 1900 and 1910 census and backward to the 1850 and 1860 census records. These linkages allow demographic data to be collected not only for the veteran, but also his childhood and adult families. Newer samples of veterans have also been linked to the 1870 and 1880 censuses and the 1920 and 1930 censuses that have become publicly available since the collection project began. See Table 1 for details on sample linkages.

A database with such rich information could not be created for Confederate soldiers because they were not eligible for the UA pension. Although the Confederate states later provided pensions for their veterans, this pension information does not contain detailed medical information. In addition, because the Confederate states must pay for pensions out of state funds (UA pensions were funded by tariffs), the pensions covered fewer men.

USCT

The most important addition to the project has been the collection of black veterans from the USCT. These were conducted in two waves, an initial sample of

Table 1. Linkage rates by sample and record type (%).

	PENSION	CMSR	CMR	SURG CERT*	CENSUS
UNION ARMY ($n = 39,340$)	68.3	99.2	57.3	72.4	65.0
USCT ($n = 6,141$)	50.2	97.2	48.9	63.3	45.6
EXPANDED USCT ($n = 15,071$)	55.8	99.6	51.4	65.5	51.3
URBAN: BOSTON ($n = 1,682$)	57.1	95.8	49.1	68.2	65.9
URBAN: CHICAGO ($n = 1,589$)	50.9	98.6	53.3	73.7	57.8
URBAN: PHILADELPHIA ($n = 2,922$)	53.0	99.4	52.5	71.6	64.6
URBAN: NEW YORK ($n = 5,017$)	41.8	98.9	36.6	65.0	49.3
URBAN: BALTIMORE ($n = 1,414$)	56.5	98.7	34.6	73.3	66.3
OLDEST OLD ($n = 2,034$)**	100	95.7	48.1	95.7	99.8
ANDERSONVILLE ($n = 1,049$)***	95.4	98.3	53.5	97.4	99.0

*Linkage to Surg Certs is contingent on having a pension.
**Only veterans who lived to age 95+, as confirmed by pension, were accepted into the sample.
***Only veterans who were POWs at Andersonville prison and survived to 1900, as confirmed by CMSR and pension, were accepted into sample. Pension link rate is not 100% because some soldiers who survived to 1900 (as confirmed on pension index) had pensions in the custody of the VA which we were unable to access.

approximately 6,000 black soldiers and their white officers, which has recently been supplemented by an additional 15,000 black soldiers and white officers. African-American companies did not begin to be formed until midway through the war, and their wartime experiences were quite different from the white soldiers. However, the basic records and collection procedures for the black and white soldiers are essentially the same. Furthermore, the USCT have been linked to all available federal census records from 1850–1940, with the exception of the fire-destroyed 1890 records.

The legacy of slavery easily can be seen in many dimensions of the USCT sample. For instance, linkage rates to the 1850 and 1860 censuses were much lower, especially for recruits who were slaves prior to the war. In later life, black veterans faced discrimination in obtaining pension support, and their linkage rates to later census years were lower than for whites, partly because their life expectancy was lower than white veterans. The USCT collection is an invaluable source of data for studying the lives of black veterans in the decades following the war.

Urban veterans supplement

The Urban sample is a stand-alone oversampling of Civil War veterans who enlisted in the largest U.S. cities. Drawn in proportion to city size in 1860, the sample was designed to allow researchers to examine intra-city disparities in environmental conditions and draw inferences about the impact of ward conditions on the recruits' life-cycle aging process. It is a random sample of companies with over 50% of the company members residing in one of five target cities: Boston, Chicago, Philadelphia, New York, and Baltimore. The Urban sample consists of over 12,500 soldiers, who were then located within these five cities at the ward level.

The varying quality of records available for each city necessitated different procedures for sample identification. Using primarily Dyer's Compendium of the War of the Rebellion and State Adjutant Generals reports, researchers compiled a list of companies that had more than 50% of their recruits enlisting in each target city. Researchers then extracted names and identifying information from the Regimental Books for those companies until the sample size was reached. Military, pension, medical, and census records were collected for these urban recruits using the same procedures as the main UAV sample, though census linkage rates have improved because of new electronic resources, such as *Ancestry.com*.

Andersonville POW supplement

The CMSRs and CMRs provide information on stressors faced by soldiers in the army. We know the battles they fought in and have a record of their wounds and injuries. We also know the location and length of service. Regimental histories can also be exploited to gain more information on the wartime events encountered by the veterans of a given company.

Soldiers taken as POWs were under conditions of extreme stress, including (in some cases) long periods of malnutrition and even starvation. A recent supplement to the data includes the records of 1,000 survivors of the Andersonville POW Confederate prison. Andersonville was the most notorious Confederate prison, and the men who passed through there in our sample had a mortality rate of 40%. In 2007, our researchers began collecting data on survivors of Andersonville, using the index developed by the National Park Service (Costa 2012).

The Andersonville sample is drawn from those recruits who survived to 1900. These survivors were linked to their CMSR, CMR, pension records, and the eight federal census years with surviving manuscripts between 1850 and 1930. Of the 1,000 veterans in the sample, 197 are linked to their siblings who were soldiers, whose complete data are also collected and will be included in a future data release.

Oldest-old supplement

Studying extreme longevity often requires drawing individuals from populations large enough to generate a sufficient number of very old members. Given nineteenth-century life expectancy, even the large UA collections have a relatively small number of members reaching very old ages. The oldest-old is an over-sampling of 1,700 individuals with an age of death confirmed to be at least 95 years old that will soon be available online.

To reach the target sample of 1,700 oldest-old, a list of nearly 6,000 potential nonagenarians was compiled from many sources, including gravestone databases, obituaries, newspaper accounts, veterans associations, and the 1930 and 1940 censuses. Death dates were confirmed from the pension files, and the veterans were linked to all the census years from 1850 to 1940.

The oldest-old supplement contains the same records present in the other Veteran's collections, but often additional records are available in later years, including physician affidavits and home visits by the Veterans Administration representatives. These records contain information on care arrangements, disabilities, and cognitive limitations not generally present for younger veterans.

HUE data

For more than a decade, the *Early Indicators* team has been collecting data on the public health environment of the following major U.S. cities: Baltimore, Boston, Chicago, Cincinnati, New York (Brooklyn), and Philadelphia, from 1830 through 1930. This collection of geospatial data was drawn from thousands of maps and published data tables, and is now publicly available as the HUE data and as ICPSR 35617. Detailed changes in ward boundaries, street layouts, and other built environment were meticulously traced over a century. No comparable street-level data for studying U.S. urban history is available elsewhere.

Researchers selected the HUE study cities and variables in order to best analyze the effects of intra-urban health disparities and public health interventions on individual mortality and longevity as observed through the UA and USCT cohorts. The HUE dataset includes ward boundary changes, street networks, and ward-level data on disease, mortality, crime, water, sanitation, commerce, industry, public works, property values, and many other variables reported by municipal departments for each study city. Researchers scoured archives in each of the seven cities, as well as obtained ward-level maps from the Library of Congress.

In sum, these materials constitute a framework that allows researchers to create accurate historical spatial and tabular data and perform geospatial and statistical analysis at many scales, giving researchers a deeper look into life in the rapidly changing cities of the nineteenth century. The HUE data by itself provides a fascinating and dynamic picture of urban American history, but even more exciting is the ability to link the UA collections to the HUE data.

Veterans who resided in large cities and applied for the pension provided their street address, which allows us to precisely pinpoint the urban environment the veterans were living in. All soldiers in our target cities have been geo-located to their street addresses and can be integrated into the HUE dataset. Additionally, all locations have been assigned county codes, enabling researchers to track the mobility of veterans over time.

New and ongoing collections

Sadly, Robert Fogel died in June 2013. Leadership of the *Early Indicators* project then passed to Dora Costa at UCLA. In 2014, the National Institute on Aging awarded a five-year grant extension under a slightly different name, *Early Indicators, Intergenerational Processes, and Aging*. The change in the grant's name reflected the new push to create datasets appropriate for studying human development and aging within an intergenerational framework.

Veterans' Children's Census (VCC) collection

The *Early Indicators* data provide a mechanism for studying intergenerational processes that is unavailable in modern data. We can study the aging and mortality not only of the veterans but also of their children and their children's households. The VCC collects all possible data on veterans' children and spouses from available census records from 1850–1940 and (where available) from vital records. Because of the difficulty of obtaining later-life linkages for children of soldiers who died in the war, the VCC collection is limited to those veterans who survived until 1900.

The VCC consists of three new scientific projects, each of which collects data on children and spouses of veterans. The collection consists of three parts: (1) 1,882 POWs who survived to 1900; (2) 8,500 white non-POWs; and (3) 4,500 African-American soldiers from the USCT. These collections come at considerable cost, since it takes (on average) about five hours of searching genealogical databases and inputting the results in order to collect the descendant data for a single UA veteran. Even with online search tools, about 70,000 person-hours of researcher time will be necessary to collect the data.

The aim of the VCC database is to facilitate the study of the causal mechanisms of intergenerational transmission of health, including the transmission of the effects of stressful events from parents to children. The data also create a valuable new object of study: the aging and longevity of women. In historical intergenerational databases, women are often neglected because it can be hard to track them once they marry and change their surnames. However, the initial findings from the VCC are that about 70% of veterans' daughters are linked to at least one census outside of their father's household, even when we do not have a married name from the pension record. Similarly, a high rate of linkage to death records is possible for the

daughters of the veterans (though linkage to death certificates is a function of location and year of death).

Intergenerationally-linked aging sample (ILAS)

Although the UA sample is broadly representative of the general population because of the high rate of participation in the Army, it is still useful to utilize data that are not confined to veterans. The ILAS data were originally collected by economic historian Clayne Pope from printed family histories and genealogies from across the United States. The sample covers 39 family lineages consisting of about 15,000 households over time. The sample includes intergenerationally-linked records of 118,162 males and females born between 1577 and 1983. About 27% of the individuals in ILAS are currently linked to at least one of the federal censuses between 1850 and 1910.

The ILAS project will explore the feasibility of linking the sample individuals to the 1920–1940 censuses and recollecting the 1850–1910 census records, which were originally obtained without the assistance of modern online search capabilities. It will also merge ecological and macroeconomic variables available from the censuses and other sources. The ILAS is a valuable companion to the UAV because it contains information from the same cohort as the UAV as well as cohorts born before and after the UA cohort. It also includes important demographic groups the UAV collections do not, namely women and Southerners.

Research possibilities

In addition to describing the multiple data sources and new research tools that have become available, this article reviews a few of the exciting research questions that can be illuminated with the *Early Indicators* data collections, including social gradients in health; racial inequality; social network dynamics; mobility in time and space; the long-term economic, demographic, and health effects of local environments and public health interventions; and the intergenerational transmission of health, well-being, and longevity. We detail a few of the important demographic topics that can be addressed with the *Early Indicators* data.

Racial inequality

The *Early Indicators* data sources provide a unique resource for studying racial disparities in health and mortality. A special collection of over 21,000 black soldiers from the USCT allows direct comparison with white veterans on a number of health-related dimensions. The life expectancy of black and white soldiers provides an indicator of hardships suffered by African Americans in the decades both before and after the Civil War. Of soldiers surviving their wartime service, only 28.1% of black soldiers survive to appear in either the pension rolls or census records in 1900, compared to 44.1% of white soldiers. As an illustration of this disparity,Table 2 shows 10-year mortality rates, by age, for white and black veterans alive in 1900 with a known age and death date (an expanded analysis would account for those censored cases where post-1900 data are available, but no death date exists). These basic results can be explored further with the demographic data from the 1900 and 1910 censuses and from the pension files, which also contain detailed medical examinations.

The *Early Indicators* data also contain detailed information from medical exams conducted to determine eligibility for and level of pension support at several points in time over the life course. Because the data from the USCT was recently expanded to include 21,000 soldiers, sample sizes are now sufficient to study racial differences in both common and uncommon conditions.

The pension files also provide a view on the social and institutional consequences of racial prejudice in nineteenth-century America. Wilson (2010), for instance, studied the impact of informal liberalization in the pension system on black and white veterans. Long before Congress officially acted to liberalize pension eligibility in 1890, veterans' groups were exerting political influence to increase dramatically the numbers of people being covered by the system. This informal liberalization applied primarily to whites. Costa (2010) found that the responsiveness of retirement decisions and living arrangements to pension income was much higher for black veterans than for whites, indicating the harsh alternative black veterans faced without that income.

Socioeconomic gradients

The data contain several indicators of socioeconomic status and background, including household wealth, literacy, occupation, and place of residence. The health impacts of a life full of hard labor, for instance, can be profound. Occupations indicate important differences in socioeconomic background, and the occupations of UA soldiers can be tracked, in many cases, from their early life to death. Pension income also raised the economic welfare of many, particularly black veterans. By 1907, the

Table 2. 10-year mortality rates by race and age, 1900–1910.

Age in 1900	Whites (UA)	Blacks (USCT)
50–54	16.6%	26.5%
55–59	22.6%	33.2%
60–64	31.4%	40.0%
65–69	44.6%	53.4%
70–74	60.2%	67.5%

pension becomes primarily aged-based, and it is possible to examine the effect of income on later-life mortality while controlling for the presence of chronic health conditions (Eli 2015). Recent work (Salisbury 2014) has extended the analysis of income to widows of veterans.

Local urban environments

The *Early Indicators* collections provide an unprecedented perspective on the long-term effects of local environments on health and longevity. Recently expanded oversampling of UA recruits from major cities allows for comparisons across cities and, more important, for analysis at the ward level. As noted above, researchers have spent many years geocoding historical maps so that recruits can be placed in a physical space that contains variables such as crime, disease, employment, government, municipal, property, and vital statistics. Large-N studies that link individuals to their local neighborhood and public health data, such as infant mortality or infectious disease death rates, as well as track people over time, are not available from other historical or modern data sources. The HUE have also been used by themselves (without the veterans) to study long-run neighborhood trends for a large number of variables (Costa and Kahn 2015; Shertzer 2015).

A useful feature of studying these urban environments is that scholars can examine how social gradients affected health in an era where access to quality healthcare was not a confounding factor. Simply put, no one (rich or poor) had access to effective medical treatments for most conditions. Thus, the privileged and well-to-do, in contrast to today, faced many of the same health challenges as the poor. More privileged classes were able, however, to locate themselves in cleaner, less congested environments. The *Early Indicators* collections allow for studying the impact of local environments at several points in time and space.

The HUE data allow the analysis of the local environment on health outcomes. These data contain a variety of health indicators, including specific diseases and overall mortality. As an example, Figure 1 shows the locations of veterans living in Philadelphia and Baltimore in 1890. Colors on the map indicate the crude mortality rate of the ward, with red being the highest quintile and blue the lowest. 1890 is the year that many soldiers entered the pension system due to the liberalization of the law governing pension eligibility.

Acute stress

A potential key indicator of later health and mortality is acute stress. Almost all soldiers in the UA experienced a

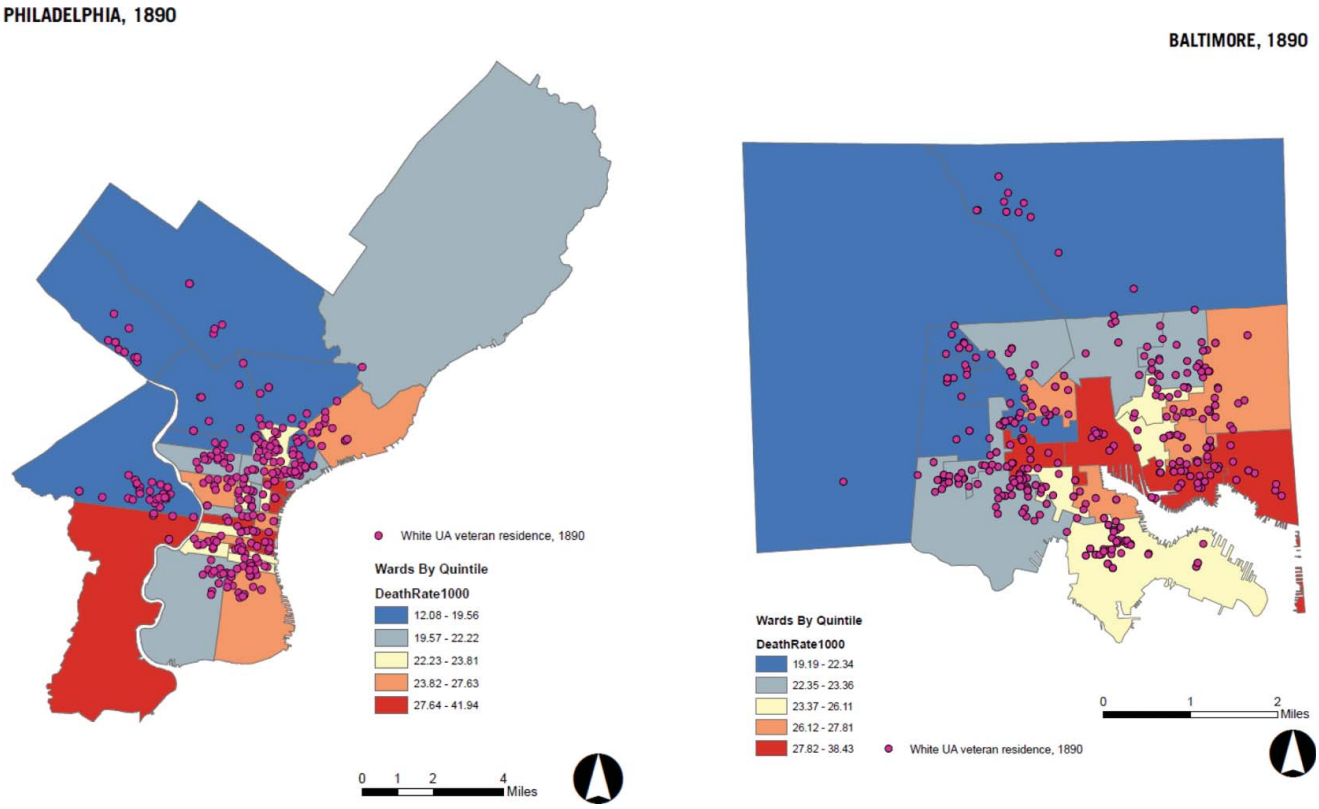

Figure 1. Veterans' residences in 1890, by ward crude mortality quintile.

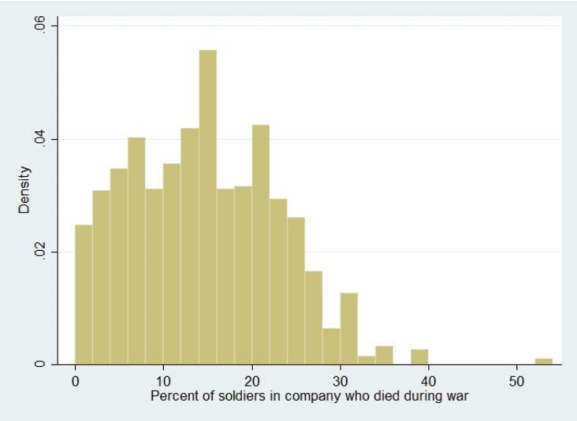

Figure 2. Company-Level mortality during the Civil War.

measure of stress, but recruits varied significantly in their experiences. The military data files contain records of which battles were experienced by recruits as well as their POW history during the war.

A simple indicator of stress is the mortality rate experienced at the company level. The distribution of company-level mortality is shown in Figure 2, which indicates the broad range of mortality risk that companies experienced. This variation can be exploited as a measure of wartime stress. Researchers can also use information in the military files to construct measures of stress. Variables that researchers might use for such an index include the battles fought, length and time of service, injuries, hospitalization, and disease at the company level. Company-level mortality has been shown to have significant effects on rates of chronic disease (Pizzaro, Silver, and Prause 2006) and on later-life mortality (Costa and Kahn 2010).

The intense stress of serving in POW camps can be studied with the Andersonville supplement discussed above. Costa and Kahn (2007a) studied the survival of POWs in the camps, finding that the presence of friends significantly increased the rate of survival. Costa (2012) has also examined the later life mortality of these same POWs. She found that older age mortality is significantly linked to the age of imprisonment. Younger soldiers at Andersonville (age <30) faced higher old-age mortality than their peers, but those imprisoned after age 30 have lower mortality in old age.

Early life risk factors

Above, we emphasized the role that acute stress during wartime had on the later life health and economic outcomes. Part of the stress experienced by soldiers was serious wounds and injuries, the consequences of which could follow soldiers in many areas of their lives. For instance, Lee (2008) found that wartime wounds and illnesses significantly diminished the veterans' geographic mobility after the war. On the health front, Wilson (2003) found that soldiers hospitalized for infectious disease during the war had significantly higher rates of chronic respiratory disease later in life.

Other scholars have used the *Early Indicators* data to link pre-war factors to later life outcomes. For example, Dejun Su (2009) found that risk exposures in early life, including season of birth, country of origin, residential region, city size, and height at enlistment influence mortality risk many decades after the war. Sok Chul Hong (2011) looked at a specific risk factor (i.e., exposure to malaria) to study the health of soldiers during the war. Those soldiers with high exposure to malaria were significantly shorter at enlistment due to malnutrition and were more susceptible to infections during the war.

One of the earliest uses of UA data was to examine the determinants of height, which is a proxy for the cumulative impact of infectious disease and nutrition during childhood. This line of research on U.S. soldiers was sparked by Robert Margo and Richard Steckel (1983), who identified the puzzling "antebellum paradox" of declining height in a period of economic growth (see also Costa 1993; Costa and Steckel 1997; Haines, Craig, and Weiss 2000). Sven Wilson and Clayne Pope (2003) found significant effects of urbanization, socioeconomic status, occupation, and migration history on the adult height of soldiers in the UA collection. More recently, Matthias Zehetmayer (2011) has found that height is positively correlated with proximity to protein-rich nutrients during childhood and with geographic mobility. Height can still be fruitfully explored with new data on African-American troops and with the augmented collection of urban recruits, where it can be combined with the wide array of ecological variables present in the HUE data.

Social networks

The *Early Indicators* collections have linked individuals in ways that are uncommon in modern datasets. Recruits were sampled as whole companies (usually 100–150 men), and companies were typically drawn heavily from a common location. Costa and Kahn (2010) have explored how these social networks affected the strength and cohesiveness of military units, and new research is examining how these social networks can persist for many decades by following the residential clustering of company members in later life (Costa et al. 2016).

Because of the extensive geocoding of the urban data, often to the street-level address, analysis of social networks can be facilitated by a geo-spatial analysis. Figure 3 shows a close-up of a Chicago neighborhood, for example, with two veterans living around the corner from one

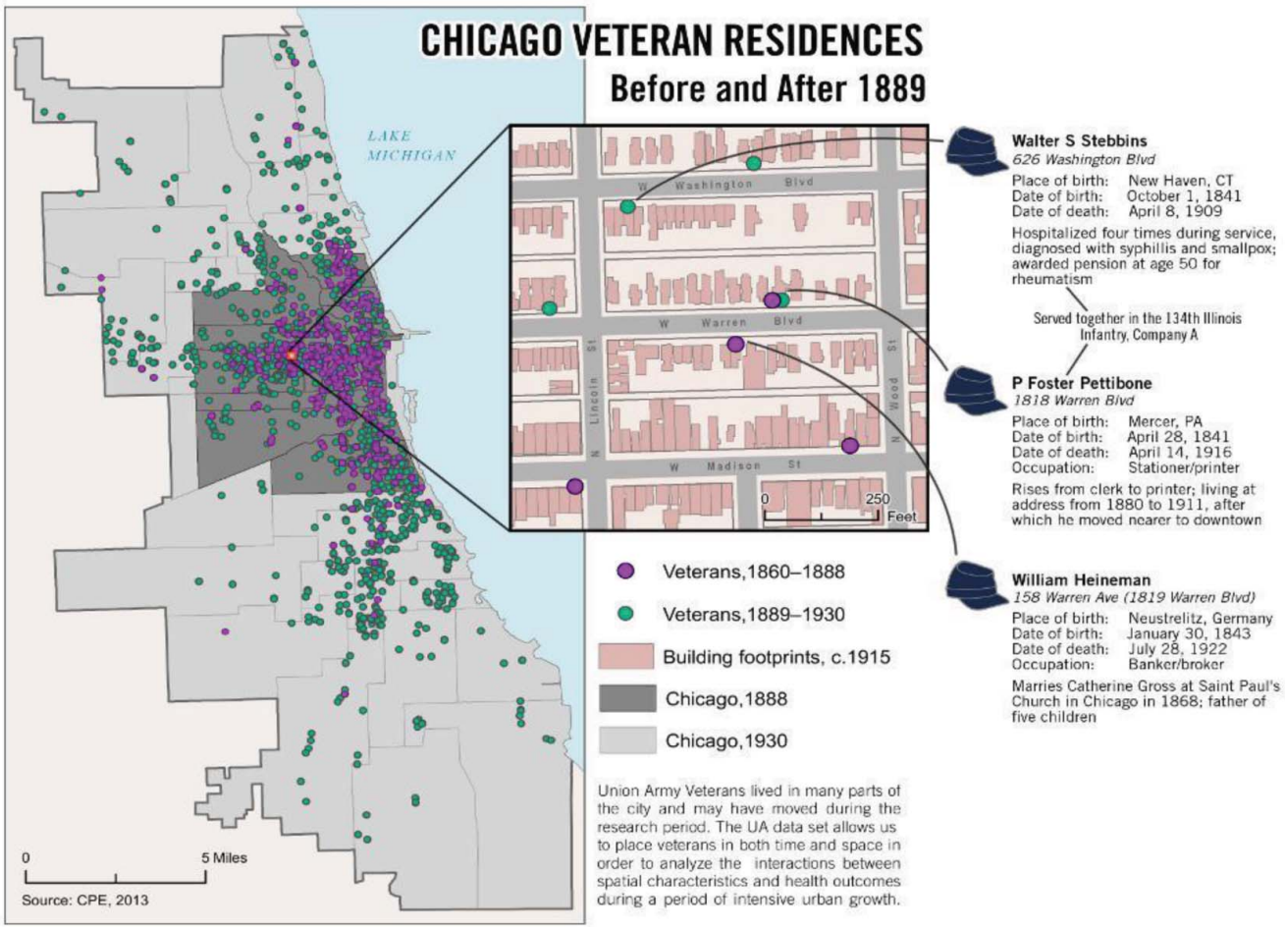

Figure 3. Veterans' residential patterns over time, Chicago.

another who served in the same Illinois company many years previously.

Family ties are also an important part of the veterans' network that can be exploited in the collection. Sizable pieces of the veterans' lives are obtained from linkage to the U.S. federal censuses from 1850–1930, allowing information about both childhood family ties and living arrangements in later life. Socioeconomic information at the family level can be used to study the family-level effects on life outcomes. Census records could be used to construct neighborhood level variables such as ethnic concentration, occupation distributions, or wealth inequality.

Elderly living arrangements

Census collections allow researchers to study aging in the context of the families and households in which people live (Costa 1997; Ruggles 2009). By 1900, most veterans were getting pension support, and they were (roughly) in the age interval 55–65, an ideal point to begin studying various aspects of the aging process related to living arrangements.

Table 3 gives one perspective on elderly living arrangements for UAV. The census enumerator indicated what member of each household was the "head," and each additional household member was identified by their relationship to the household head. As can be seen from Table 3, most veterans are listed as head, a percentage that declines gradually over the next two decades.

The most striking feature of this preliminary analysis is the similarity between black and white living arrangements of elderly soldiers. In both 1900 and 1910, blacks and whites were equally likely to be household heads.

Table 3. Living arrangements among veterans, by race and year.

	1900		1910	
Relationship to head	Whites	Blacks	Whites	Blacks
Head	88.7%	88.8%	82.0%	83.3%
Parent	2.5%	1.8%	8.2%	3.6%
Spouse or relative	2.0%	1.9%	2.0%	3.4%
Non-relative	4.5%	6.7%	5.7%	7.6%
Alive but not linked to census	2.3%	0.8%	2.2%	2.2%
N	12,737	4,201	7,567	2,600
Average age (census age)	60.3	59.8	69.0	68.9
Age range	50–79		60–89	

The likelihood of living as a parent of the household head is, in 1910, 8.2% for whites and 3.6% for blacks. It is likely the case that black veterans were more likely to have adult children who were less likely to live independently, possibly for economic reasons.

This analysis does not capture the rate at which white and black veterans may have lived with adult children as household members (in other words, what percentage of the veterans who were household heads had adult children or grandchildren living with them). Future work includes bringing more of this detail into the analysis, as well as information on the veterans' health, disability, and pension support. Of course, bringing in national comparison data from the IPUMS data will allow comparisons between veteran households and the general population.

Chronic disease epidemiology

Research on the secular decline of chronic conditions and disabilities has been published with earlier versions of the data that has reshaped how scholars think about the epidemiological transition (Costa 2000; Fogel and Costa 1997; Wilson et al. 2005). In addition, the data collection remains a rich source of information for studying specific chronic diseases. Past work includes an examination of arteriosclerosis (Costa, Helmchen, and Wilson 2007), respiratory disease (Wilson 2003), and hernias (Song and Nguyen 2003). Andrew Noymer (2009) used the UA data to test whether tuberculosis was a risk factor for influenza epidemic, thereby reducing the prevalence of tuberculosis in the population.

These examples use only a small subset of the available data with which to study chronic disease epidemiology across the life course. Researchers on the *Early Indicators* team have worked for years to develop a system of standardizing the findings from the medical examinations that allow for detailed study of the epidemiology of dozens of chronic conditions in later life. Of course, an enormous advantage of the UA collections is that risk factors for those conditions can be identified back to the early life experiences of the veterans.

There have been studies looking at the overall burden of chronic diseases and disabilities using the UA collection. Costa (2000, 2002) and Fogel and Costa (1997) documented and explored the widespread decline in chronic disease rates and the associated decline in functional limitations. Costa (2000) emphasized the role that the decline in manual labor and exposure to infectious disease as primary contributors to the decline in condition prevalence over a wide number of conditions. Wilson and colleagues (2005) emphasized the importance of the decline in non-fatal conditions that in the nineteenth and early twentieth century were highly prevalent and disabling but today seldom lead to disability; these include hernias, varicose veins, and a large number of gastrointestinal disorders.

Intergenerational forces

As noted above, family connections are an important aspect of the data collections. A recent renewal of funding by the National Institute on Aging allows the collection of data on the children (and the households of those children). The *Early Indicators* team has already begun the process of linking the census and death information for the children of the UAV, both white and black. The new collection will also significantly increase the amount of data available on women, as the project collects that same data for both the sons and daughters of the veterans. New electronic search tools, such as those available from *Ancestry.com*, also yield higher find rates for women than have been possible in the past.

We will thus be able to follow individuals across three generations and study the factors in the lives of the veterans that affect life outcomes into the next generation. For instance, from the 1850 and 1860 censuses, it is frequently possible to find the occupations of the soldiers' fathers, and the new data will give the occupations of their children all the way through the 1940 census. We know of no other data source that allows for the systematic study of multigenerational social mobility and how that mobility intersects with health and longevity, not to mention other social factors such as race or immigration history.

As an illustration, Table 4 combines data from the POW sample and a small VCC pilot done on white soldiers in the main sample. The table shows the occupations of veterans' sons as a function of age and birth order. In this table, first-born sons in their twenties are less likely to start their adult lives as laborers and more likely to be in professions or skilled trades, though that gap narrows as the brothers age. Moreover, the younger

Table 4. Occupation of veterans' sons, by birth order.

	Age in 1900: 20–29			Age in 1900: 30–39		
Occupation in 1900	1st son (%)	2nd son (%)	3rd son (%)	1st son (%)	2nd son (%)	3rd son (%)
Laborer	33	41	43	29	25	29
Farmer (owner/ operator)	21	18	23	20	29	31
Trade/craft/clerical	40	36	32	40	40	36
Manager/ professional	6	5	3	11	6	4
Total	100	100	100	100	100	100
N	349	253	160	159	164	107

brothers appear to move from being laborers to farmers as they age, which does not happen for first-born sons. Sample sizes in this analysis are much too small to draw any conclusions, but one can envision using the ultimate VCC sample, once it has been collected, to study a host of demographic and economic factors related to social mobility in an intergenerational context.

Persistent challenges and continuing progress

The UA collections are a vast and complex set of data covering a dynamic century of economic growth, industrialization, urbanization, and technological innovation. These features make these data both immensely valuable for research and immensely challenging to understand and use appropriately. Any researcher using the collections must be aware of the significant challenges involved in avoiding the many biases that can arise when using the data.

Understanding and accounting for bias

A thorough discussion of potential biases in the UA collections can be found on the *uadata.org* website and, to a large extent, in the many published papers using the data. In this section, we highlight (in broad terms) some key issues that researchers must consider when using the UA collections.

In an ideal situation, data on all sample members are generated by the same data-generating process with parameters that depend only on factors that all sample individuals are at risk to experience at each point in time. The UA collections deviate from this ideal in numerous ways. We highlight some of these in the following paragraphs.

Bias due to event-generated data

As noted above, the data collection captures information across the soldier's life course. The existence of data at one point in time often depends on events that happen at a *future* date, which creates significant survival bias. The most ubiquitous source of such survival bias is caused by the event of applying for a UA pension. The pension application consists of a thick file of documents that describe life for many decades *before* the pension, including experiences during the war. Thus, some data in the collections during early life, during the war, and after the war, exist because the veteran lived to enter the pension system, often much later in life.

The obvious complication in conducting analysis is that the data on two identical soldiers at a point in time (say 1870) will be very different if one of those soldiers survived to apply for a pension at a later date (say 1880), but the other did not. The pension also contains demographic and economic information, such as marriage and family data, and occupation for the period covered between the war and the time of pension application. If the veteran dies without applying for this pension, this information is unavailable, even though he may have lived for many decades beyond the end of the war.

One particularly problematic source of bias is when information from the pension files is used to populate data fields describing the wartime experience, such as claims of injuries or illnesses. In the main UAV file, data that come from military files and data coming from later pension applications are not distinguished. Special files have been created that give the information from CMSRs and CMRs without using any information from the pension application. Researchers needing an unbiased view of the wartime experience should utilize these important auxiliary files.

Surviving to participate in the pension system increases the likelihood of linking the veteran to the 1850 and 1860 censuses because information obtained from the pension files and from the 1900 and 1910 censuses is very useful in obtaining the census records from the veteran's childhood. Thus, an unbiased analysis of early life effects on later life outcomes is best conducted by conditioning on survival to a point in time where most veterans would have entered the pension system. If one wanted to address, for instance, the effect of early life environment on, say, survival during the war, the researcher must account for the fact that the early life variables are much less likely to exist in the data for soldiers who died during the war and, therefore, never obtained a pension. Thus, the collection methods used to make linkages to the census impose serious limitations on how the early life census data can be used for research.

Another significant event that further complicates the nature of survival bias was the liberalization of the pension system in 1890. Prior to that point, veterans must make an argument that their disability was "war-related." Under the 1890 law, any disability (regardless of cause) that limited the capacity to perform manual labor was eligible for pension support. This caused an immediate and significant increase in pension enrollment, which in turn generated a host of retrospective data of the kind we have been discussing. One important complication of this change for research is that any time series of disease or disability prevalence calculated from the medical data in the pension files must account for the difference in eligibility (and hence existence of the data) that occurred in 1890.

Variation and evolution in census linkage rates

Much of the important demographic information on the veterans' lives comes from the census records. An

unavoidable problem in using these data is that linkage to the census is not random; it is correlated with factors that may, in turn, be correlated with the life outcome of interest. For example, urban environments were known to be much less healthy than rural ones, but soldiers in urban environments are also much harder to identify in the census because census linkage is difficult in urban areas. Place of birth is another variable that may be correlated with both the linkage rate and the outcome of interest.

The past decade has seen an explosion of new Internet-based search tools available through sites such as *Ancestry.com*. These tools are now the primary method used to link veterans to both census records and vital records. This has led to much higher linkage rates for the later collections, namely the USCT, Urban, Andersonville, and Oldest-Old samples. This increased linkage rate is very good news for users of those samples but a problem for researchers comparing those soldiers with the main UA collection, which was collected without those tools.

For all data collection after the original UA sample, researchers used *Ancestry.com* to link living veterans to all census decades available. Linkage rates are high for veterans who survived to enter the pension system because information from the Civil War pension record is used to identify soldiers on the census. This information includes, but is not limited to, alternate name spellings and aliases, birthdate and place, death date and place, residences, family information (spouse, children, and sometimes parents and siblings), and occupation. A search begins with basic information, such as name, birthdate, birthplace, and residence location. Other pension information is used to determine if an individual found in the search results is the veteran. Each census decade contains slightly different information, so the exact information used to search and confirm a census link varies, but it is all information found in the pension record. Linkage rates are high and the matches robust because so much is known about the veteran and his family from the pension record.

Many researchers currently are exploring mechanized census linking. Our preliminary work with mechanized census linkage has yielded lower linkage rates than those obtained by our inputters and suggested a bias against finding men who were not upwardly mobile relative to census linkage by highly-trained inputters. Martha Bailey (2016) has compared the links created through clerical review by staff for her Longitudinal, Intergenerational Family Electronic Microdata (LIFE-M) project to different machine matching schemes. In her preliminary results, her matching rates were similar to those of some of the machine algorithms, but the machine algorithms had false positive rates of up to 60%. *Early Indicators* avoids almost all these false positives by employing highly-trained research assistants who use pension information for searching and confirming matches.

Historical change

The pension data cover decades of time in which advances in technology, medical knowledge, and public health were occurring. These changes are reflected in the medical records found from the pension files. A typical physical examination record from 1870 looks vastly different from a typical examination record 50 years later in 1920. The increase in knowledge and professionalization of the examiners is readily apparent from even a casual observation. Such a transformation does not necessarily cause bias, but researchers must be careful in constructing time series of health outcomes from the pension files.

Administrative racial prejudice

An exciting new feature of the UA collections is the large, new collection of records on African-American soldiers in the USCT. One of the first things many researchers will want to do is make comparisons between white and black soldiers. This, indeed, is one of the primary motivating goals behind the collection of the USCT data, but it may be biased by underlying discrimination faced by black soldiers both during the war and in the pension system.

Wilson (2010) provided a detailed discussion of how prejudice in the pension system was very effective at keeping black soldiers' rate of participation low. Part of this lower participation is due to the fact that blacks were less likely to be injured in battles than white soldiers, but a large part of the differential is due to discrimination. Figure 4 shows estimated enrollment in the pension system by race and by wartime medical history. Controlling for evidence in the military records, by the time the 1890 Act was passed, whites had significantly higher enrollment than blacks in each category. Pension enrollment rates for those with a war wound were about 60% for whites, but only 35% for blacks. For those with an illness but no wound, whites were at 43%, and blacks were at 10%. For those without any illness or wound, whites were at 22%, and blacks were at 8%.

In the post-Reconstruction period, it is easy to see from Figure 4 the informal liberalization of the pension system that was occurring between 1879 and 1889 prior to the formal liberalization in 1890. Starting with the passage of the 1879 Arrears Act, which allowed back payments of pension support, this was a period of intense political activism on behalf of veterans. Blacks did not share equally from this liberalization. Their claims of war-related disabilities were rejected at higher rates, and they had a harder time gathering documentation that allowed

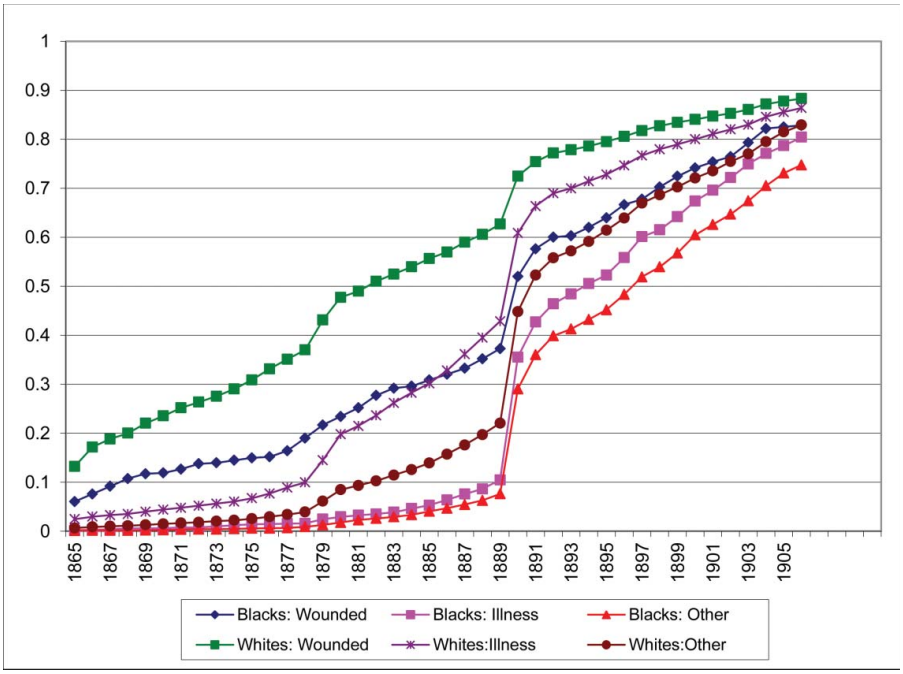

Figure 4. Pension enrollment proportions, by race and wartime medical history. *Source*: Wilson (2010).

them to verify their identities and their service. This was due to their status as slaves prior to service and to treatment during the war.

In the *Early Indicators* data, 18.7% of the black sample died from disease while in service, compared to only 9.6% of whites. Yet, black troops were less likely to be sent to the hospital for illness than white troops. Scholars have argued that white medical officers "accused blacks of feigning sickness in much the same way that masters and overseers accused slaves of shirking work. They mistreated, abused, overworked, or neglected such soldiers, thereby contributing to further deterioration of their health" (Berlin, Reidy, and Rowland 1998, 636). Margaret Humphreys (2008) noted that black regiments were also understaffed compared to white regiments. The short-term impact of this discrimination was higher disease mortality during the war. The long-term impact was that blacks lacked documentation of their wartime illnesses, leading to a greater difficulty in claiming pension support in the period of informal liberalization.

In sum, the UA collections allow for detailed comparisons of the health of aging veterans, both black and white. The blatant discrimination in the pension system is a fact that researchers must grapple with in making black-white comparisons using medical data from the pension files.

User services

Early Indicators researchers have put extensive work into easing the challenging task of using the UA collections, including cleaning and standardizing the data values, many of which are free-form text. As a general principle, the data inputters were trained to avoid making ad hoc corrections and standardizations. The project leadership team has tried to strike a balance between preserving all the relevant information and making individual variables as easy to use as possible. This is ongoing work in the current grant cycle. Efforts include the creation of "derived variables" that can be used easily in analysis. Data standardization is most evident in the medical data from the Surgeon Certificates, since most of that data is free-form entries from medical exams. On the *UAdata.org* website, users can download sets of common standardized variables. These user-friendly variables are available for several categories as "Data 101" files. There is no shortcut to carefully studying the accompanying documentation, including sections on sample design and on the bias issues discussed above, but there are tools available on *UAdata.org* that can considerably flatten the learning curve.

First, although the data collections and documentation can be downloaded in bulk for users accustomed to working with the data, the new data extraction system can be a valuable tool for the new researcher. The extraction system is organized topically, allowing a user to focus on a set of variables from all data sources related to the topic of interest. For example, Figure 5 shows a subset of the results from a search on the topic of marriage, illustrating both census and pension-based variables. The columns shown in the extraction system provide the user with quick links to

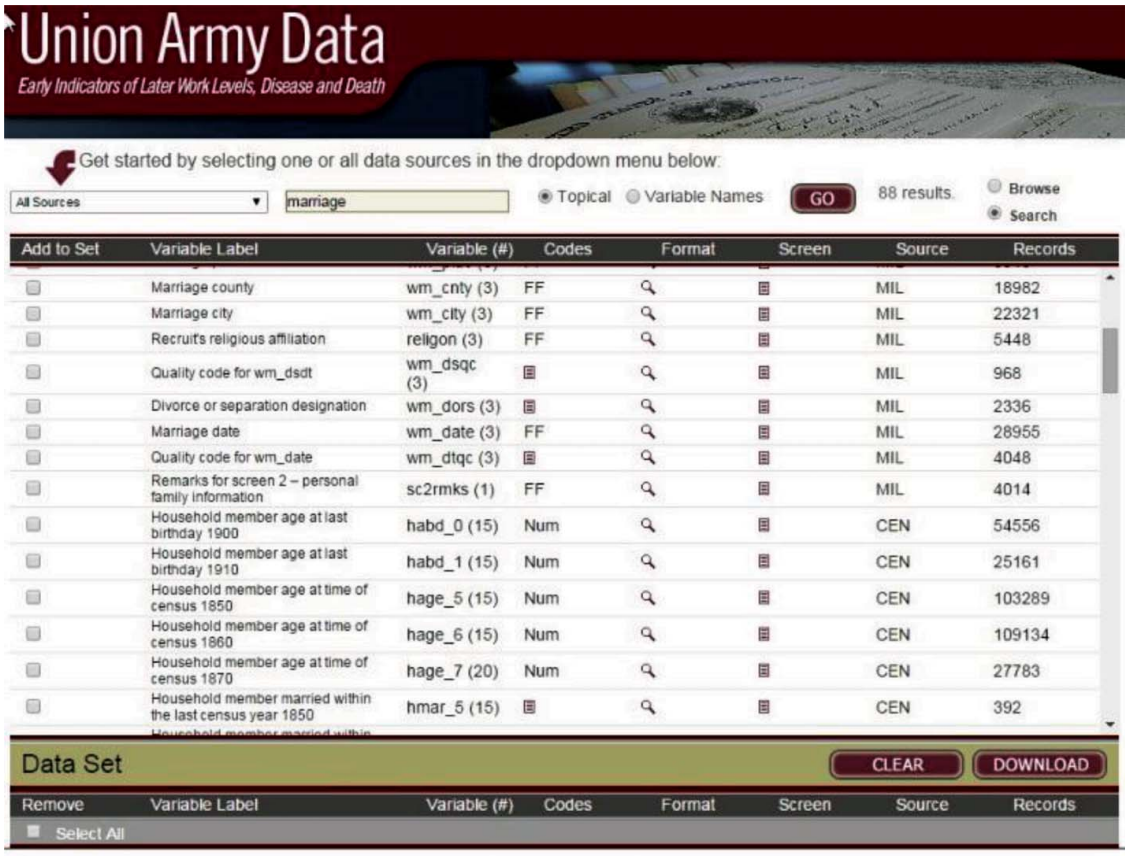

Figure 5. Union Army data searchable extraction system.

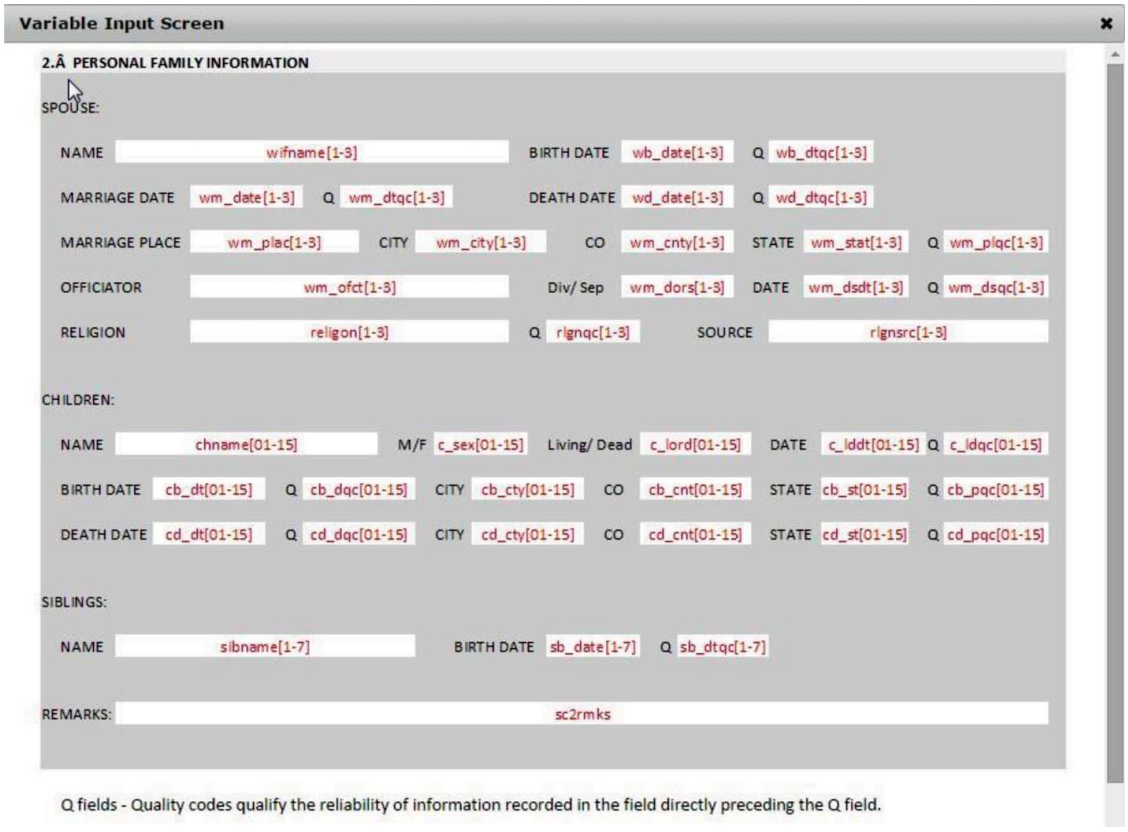

Figure 6. Example of data inputting screen.

the codes and formatting used for each variable. Each variable is also linked to the data collection screen from which it comes. This is an invaluable tool for learning how variables relate to each other, and the user can see, in this example, the marriage information available on Screen 2.A., as well as a variety of other family information, including the variable names.

The data extraction system is a helpful tool for users to quickly obtain needed variables. After selecting the variables they want, users can indicate on the export page which collection they want to draw from (Whites, Blacks, Urban, etc.), and how they want the data formatted (Excel, SAS, Stata, etc.). Each export comes with a customized set of online documentation that provides general background information as well as documentation that is unique to the exported variables.

Despite its name, the primary function of the extraction system is not extraction, but teaching. The topic-based approach to the data, which can be browsed or searched, is designed to help users understand the collections and to ease the path of entry into these complex data collections. For example, Figure 6 shows an example of a data inputting screen for personal family information with variable names highlighted in the fields; these are available for all variables and screens. In addition, the website provides a large set of topical user guides that provide the user with extensive detail on the variables and the sample design and collection methods used to obtain them. The extraction system is currently available for the UAV and original USCT samples, with all other datasets to be added soon. Personal assistance from the *Early Indicators* team is also readily available, and the staff will respond to queries as soon as possible.

Conclusion

The collection of the UA data began nearly 30 years ago. Since that time, these soldiers have been contributing to our understanding of health, mortality, economic mobility, and a variety of scientific questions across many disciplines. By this measure, the UA collections are old news, but we have tried to illustrate with this article many vital questions related to the human experience that might be illuminated by the UA collections are still to be explored and answered. Many of these questions are best served by having data on a large set of individuals over their entire lifetimes. Some key forces that seem to shape human health and well-being (i.e., poverty, stress, inequality, discrimination, family dissolution, and environmental risks) can be addressed with this remarkable and growing set of life histories.

Funding

This research is funded by the National Institute on Aging (P01 AG10120).

References

Alter, G., and J. Riley. 1989. Frailty, sickness and death: Models of morbidity and mortality in historical populations. *Population Studies* 43:25–45.

Bailey, M. 2016. *LIFE-M longitudinal, intergenerational family electronic microdata*. Presented at the NBER Cohort Studies Group Presentation.

Berlin, I., J. P. Reidy, and I. S. Rowland. 1998. *Freedom's soldiers: The black military experience in the Civil War*. Cambridge, UK: Cambridge University Press.

Cain, L., and S. C. Hong. 2009. Survival in 19th century cities: The larger the city, the smaller your chances. *Explorations in Economic History* 46(4):450–63.

Canavese, P., and R. Fogel. 2009. Arthritis: Changes in its prevalence during the nineteenth and twentieth centuries. In *Health at older ages: The causes and consequences of declining disability among the elderly*, eds. D. M. Cutler and D. A. Wise, 55–74. Chicago: NBER and University of Chicago Press.

Costa, D. L. 1993. Height, wealth, and disease among the native born in the rural antebellum North. *Social Science History* 17:355–83.

Costa, D. L. 1997. Union Army pensions and elderly living arrangements. *Journal of Political Economy* 105:1269–92.

Costa, D. L. 1998. *The evolution of retirement: An American economic history, 1880–1990*. Chicago: NBER and the University of Chicago Press.

Costa, D. L. 2000. Understanding the 20th century decline in chronic conditions among older men. *Demography* 37:53–72.

Costa, D. L. 2002. Changing chronic disease rates and long-term declines in functional limitations among older men. *Demography* 39:119–38.

Costa, D. L. 2010. Pensions and retirement among black Union Army veterans. *Journal of Economic History* 70(3):567–92.

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

Costa, D. L. 2015. Health and the economy in the United States from 1750 to the present. *Journal of Economic Literature* 53 (3):503–70.

Costa, D. L., L. Helmchen, and S. E. Wilson. 2007. Race, infection, and arteriosclerosis in the past. *Proceedings of the National Academy of Sciences* 104:13219–24.

Costa, D. L., and M. E. Kahn. 2007a. Surviving Andersonville: The benefits of social networks in POW camps. *American Economic Review* 97(4):1467–87.

Costa, D. L., and M. E. Kahn. 2007b. Deserters, social norms, and migration. *Journal of Law and Economics* 50:323–53.

Costa, D. L., and M. E. Kahn. 2008. *Heroes and cowards: The social face of war*. Princeton, NJ: Princeton University Press.

Costa, D. L., and M. Kahn. 2010. Health, wartime stress, and unit cohesion: Evidence from Union Army veterans. *Demography* 47(1):45–66.

Costa, D. L., and M. E. Kahn. 2015. Declining mortality inequality within cities during the health transition.

American Economic Review (Papers and Proceedings) 105:5642–569.

Costa, D. L., M. E. Kahn, C. Roudiez, and S. E. Wilson. 2016. Persistent social networks: Civil War veterans who fought together co-locate in later life. NBER Working Paper #22397. Cambridge, MA: National Bureau of Economic Research.

Costa, D. L., and J. Lahey. 2005. Predicting older age mortality trends. *Journal of the European Economic Association* 3:487–493.

Costa, D. L., and R. Steckel. 1997. Long-term trends in health, welfare, and economic growth in the United States. In *Health and welfare during industrialization*, eds. R. H. Steckel and R. Floud, 47–90. Chicago: University of Chicago Press.

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

Eli, S. 2015. Income effects on health: Evidence from the Union Army pensions. *Journal of Economic History* 75:448–78.

Fogel, R.W. 1986. Nutrition and the decline in mortality since 1700: Some preliminary findings. In *Long-term factors in American economic growth*. Chicago: University of Chicago Press.

Fogel, R. W. 1993. New sources and new techniques for the study of secular trends in nutritional status, health, mortality, and the process of aging. *Historical Methods* 26:5–43.

Fogel, R. W., and D. L. Costa. 1997. A theory of technophysio evolution, with some implications for forecasting population, health care costs, and pension costs. *Demography* 34:49–66.

Fogel, R. W., and S. L. Engerman. 1974. *Time on the cross: The economics of American Negro slavery*. Boston: Little, Brown.

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

Gluckman, P. D., M. A. Hanson, H. G. Spencer, and P. Bateson. 2005. Environmental influences during development and their later consequences for health and disease: Implications for the interpretation of empirical studies. *Proceedings: Biological Sciences* 272:671–77.

Hacker, J. D. 2008. Economic, demographic, and anthropometric correlates of first marriage in the mid-nineteenth-century United States. *Social Science History* 32:307–45.

Haines, M. R. 1979. The use of model life tables to estimate mortality for the United States in the late nineteenth century. *Demography* 16:289–312.

Haines, M. R., L. A. Craig, and T. Weiss. 2000. *Development, health, nutrition, and mortality: The case of the "antebellum puzzle" in the United States*. NBER Working Paper No. H0130. Cambridge, MA: National Bureau of Economic Research.

Hong, S. C. 2011. Malaria and economic productivity: A longitudinal analysis of the American case. *Journal of Economic History* 71:654–71.

Humphreys, M. 2008. *Intensely human: The health of the black soldier in the American Civil War*. Baltimore: The Johns Hopkins University Press.

Komlos, J. 1987. The height and weight of West Point cadets: Dietary change in antebellum America. *Journal of Economic History* 47:897–927.

Lee, C. 1998. Long-term unemployment and retirement in early-twentieth-century America. *Journal of Economic History* 58:844–56.

Lee, C. 2001. The expected length of male retirement in the United States, 1850–1990. *Journal of Population Economics* 14:641–50.

Lee, C. 2005. Wealth accumulation and the health of Union Army Veterans, 1860–1870. *Journal of Economic History* 65 (June):352–85. Earlier version published in NBER Working Paper No. 10035.

Lee, C. 2007. Military positions and post-service occupational mobility of Union Army veterans, 1861–1880. *Explorations in Economic History* 44(October):680–98. Earlier version published in NBER Working Paper No. 12415.

Lee, C. 2008. Health, information, and migration: Geographic mobility of Union Army veterans, 1860–1880. *Journal of Economic History* 68(3):862–99.

Noymer, A. 2009. Testing the influenza-tuberculosis selective mortality hypothesis with Union Army data. *Social Science Medicine* 68(9):1599–1608.

Margo, R. A., and R. H. Steckel. 1983. Heights of native-born whites during the antebellum period. *Journal of Economic History* 43(1):167–74.

McKeown, T. 1976. *The modern rise of population*. New York: Academic Press.

McKeown, T. 1978. Fertility, mortality and cause of death: An examination of issues related to the modem rise of population. *Population Studies* 32:535–42.

Myers, G. C., V. L. Lamb, and E. M. Agree. 2003. Patterns of disability change associated with the epidemiological transition. In *Determining health expectancies*, ed. J. M. Robine, C. Jagger, C. Mathers, E. Crimmins, and R. Suzman, 59–74. Chichester, England: John Wiley & Sons.

Pizarro, J., R. C. Silver, and J. Prause. 2006. Physical and mental health costs of traumatic war experiences among Civil War veterans. *Archives of General Psychiatry* 63:193–200.

Pope, C. L. 1992. Adult mortality in America before 1900: A view from family histories. In *Strategic factors in nineteenth century American economic history*, eds. C. Goldin and H. Rockoff, 267–96. Chicago: University of Chicago Press for NBER.

Riley, J. C., and G. Alter. 1989. The epidemiologic transition and morbidity. *Annales de De'mographie Historique* 1989:199–213.

Ruggles, S. 2009. Reconsidering the Northwest European family system: Living arrangements of the aged in comparative historical perspective. *Population and Development Review* 35:249–73.

Salisbury, L. 2014. *Women's income and marriage markets in the United States: Evidence from the Civil War pension*. NBER Working Paper No. 20201. Cambridge, MA: National Bureau of Economic Research.

Shertzer, A. 2015. *Immigrant group size and political mobilization: Evidence from European migration to the United States*. NBER Working Paper No. 18829. Cambridge, MA: National Bureau of Economic Research.

Song, C., and L. L. Nguyen. 2003. The effect of hernias on labor force participation of Union Army veterans. In *Health and labor force participation over the life cycle: Evidence from the past*, ed. D. L. Costa, 253–310. Chicago: NBER and University of Chicago Press.

Su, D. 2009. Risk exposure in early life and mortality at older ages: Evidence from Union Army veterans. *Population and Development Review* 35(2):275–95.

United Nations. 1973. *The determinants and causes of population trends*. Population Studies no. 50. New York: United Nations Publications.

Verbrugge, L. M. 1984. Longer life but worsening health? Trends in health and mortality of middle-aged and older persons. *The Milbank Memorial Fund Quarterly, Health and Society* 62:475–519.

Wilson, S. E. 2003. The prevalence of chronic respiratory disease in the Industrial Era: The United States, 1895–1910. In *Health and labor force participation over the life cycle: Evidence from the past*, ed. D. L. Costa, 147–80. Chicago: NBER and University of Chicago Press.

Wilson, S. E. 2010. Prejudice and policy: Racial discrimination in the Union Army disability pension system, 1865–1906. *American Journal of Public Health* 100:S56–S65.

Wilson, S. E., J. Burton, and B. Howell. 2005. *Work and the disability transition in twentieth century America*. NBER Working Paper No. 11036. Cambridge, MA: National Bureau of Economic Research.

Wilson, S. E., and L. Nguyen. 1998. Secular trends in the determinants of disability benefits. *American Economic Review* 88: 227–231.

Wilson, S. E., and C. L. Pope. 2003. The height of Union Army recruits: Family and community influences. In *Health and labor force participation over the life cycle: Evidence from the past*, ed. D. L. Costa, 113–46. Chicago: NBER and University of Chicago Press.

Wimmer, L. 2003. Reflections on the early indicators project a partial history. In *Health and labor force participation over the life cycle: Evidence from the past*, ed. D. L. Costa, 1–10. Chicago: NBER and University of Chicago Press.

Zehetmayer, M. 2011. The continuation of the antebellum puzzle: Stature in the US, 1847–1894. *European Review of Economic History* 15(2):313–27.