

[illegible]

IPUMS Full Count Datasets of the United States Censuses of Mortality, 1850-1880

J. David Hacker, Lap Huynh, Susan H. Leonard, Matt A. Nelson, Evan Roberts, and Matthew Sobek

Abstract:

This article describes four new IPUMS datasets constructed from the 1850, 1860, 1870, and 1880 Censuses of Mortality of the United States. We discuss the creation of the datasets, the variables included in each census year, and their potential for social science research. We highlight several limitations in the data and caution users about potential biases. Finally, we illustrate the usefulness of the new data by analyzing the relationship between household wealth and child mortality in 1870. All four datasets and associated documentation are distributed for public use via the IPUMS website.

Author Information: J. David Hacker (corresponding author, hacker010@umn.edu), University of Minnesota, Department of History and Minnesota Population Center, <https://orcid.org/0000-0002-5971-955X> ; X(Twitter): @jdavidhacker612; Lap Huynh, University of Minnesota, Institute for Social Research and Data Innovation; Susan H. Leonard, University of Michigan, Inter-University Consortium for Political and Social Research; Matt A. Nelson, University of Minnesota, Institute for Social Research and Data Innovation, <https://orcid.org/0000-0002-8849-4628> , X(Twitter):@mattnelson; Evan Roberts, University of Minnesota, Program in the History of Medicine, ORCID: <https://orcid.org/0000-0001-5621-4823>, X(Twitter): @evanrobertsnz; Matthew Sobek, University of Minnesota, Institute for Social Research and Data Innovation

Introduction

Life expectancy and the risk of death are essential measures of population wellbeing. Unfortunately, the lack of individual-level data on the timing and causes of death in the United States prior to 1900 has been a significant obstacle in the study of differences in life expectancy and mortality among population subgroups, across space, and over time. In this paper we describe new IPUMS full count datasets for the Censuses of Mortality conducted by the United States in 1850, 1860, 1870, and 1880. The datasets, which together provide details on 1,829,279 persons who died in the year prior to the population censuses in these years, were constructed at the University of Minnesota in partnership with the private genealogy company Ancestry. Funding was provided by the Eunice Kennedy Shriver National Institute of Child Health and Human Development (NICHD). All datasets are available for public use on the IPUMS website (www.ipums.org).

Creation of microdata sets for the censuses of mortality varied in important ways from the full count microdata sets for the population censuses described in an accompanying paper in this journal issue. Unlike the manuscript returns for the population census, which were archived and microfilmed by the National Archives after its creation in 1934, the mortality censuses were held by a variety of state and local archives. Not all records have survived, and many had not been microfilmed prior to the beginning of this project. For each census year, we created datasets encompassing all census of mortality records that had been digitized by Ancestry, supplemented with additional surviving records that we located, photographed, or microfilmed, and entered by hand in Minnesota. The “full count” mortality datasets are therefore only complete transcriptions of surviving records. Fortunately, a significant percentage of records have survived. The datasets range from a low of approximately 85.0% of the returns in 1870 to a high of 98.0% in 1860. In addition, as we discuss below, some fields on the

mortality schedules, such as occupation in census years 1860-1880, were not captured by Ancestry.¹ For census years 1870 and 1880, we were also able to link a significant subset of the decedents to their households of origin in the corresponding population census, owing to the presence of a common household number used in the two sources. These linked data facilitate additional analyses, including more detailed examination of the social determinants of mortality.

We describe here each dataset's method of construction, variable availability, strengths and weaknesses, and potential for social science research. We also demonstrate one potential use of the datasets by analyzing the association between parental wealth and child mortality in 1870.

Background

The 1850-1880 United States censuses of mortality, which were conducted concurrently with the decennial population censuses, were meant to fill a major gap in the nation's collection of demographic data. Although a national census was carried out every ten years beginning in 1790, vital registration was left to the states. A few states and municipalities, beginning with Massachusetts in 1842, followed England's lead in establishing vital registration systems in the mid nineteenth century, but most states and municipalities made no effort to collect data on births and deaths (Vinovskis 1972, Preston and Haines 1991). In 1900, when the "national" death registration area (DRA) was first established, just 40.3% of the population of the United States lived in the 13 states or the District of Columbia where deaths were registered (Haines 1979). Correctly anticipating that a comprehensive national system of vital registration was many decades away, nineteenth-century statisticians lobbied Congress to have census enumerators collect retrospective mortality information from households while conducting the population census. A proposal to conduct a mortality census was first approved with

¹ Names of decedents have been removed in the public use versions of the datasets. Restricted-use versions with names are available for investigators with a demonstrated research need.

other revisions to the 1850 population census (Anderson 1989: 37). After enumerating each free household and slave holding, 1850 census enumerators were to inquire whether any deaths had occurred in the household or holding in the year prior to the census. Because the nominal census day was June 1, 1850, the inquiry was related to deaths between June 1, 1849, and May 31, 1850. If a death had occurred, the enumerator recorded the decedent's name, color (race), sex, age, marital status, place of birth, occupation, month of death, and cause of death on a separate schedule. Beginning in 1870, for places with an established death registration system, census enumerators entered information directly from the registration records rather than collecting the data with the population census.² This method was greatly expanded in 1880. Despite many criticisms of the results, the mortality data were analyzed by the Census Office in separate publications. These censuses represent the only nationally representative individual-level mortality data prior to the twentieth century and are therefore a potentially valuable source for demographic historians.

The criticisms of the censuses of mortality, however, were substantial, both in the historical era when the censuses were conducted and by social scientists in the twentieth century. In an era in which infectious diseases were the leading killers, and when the nation experienced high year-to-year variations in death rates, mortality and causes of death in the year prior to each census may not have been representative of surrounding years. A cholera pandemic in 1849, for example, likely resulted in more deaths from cholera and more deaths overall than in other nearby years (Vinovskis 1978). Second, cause of death data, which was reported by household members, reflected substantial reporting errors. Prior to the germ theory of disease in the 1880s, causes of death were poorly understood, even by

² As discussed below, these data, which were primarily for New Jersey, Massachusetts and 20 cities, were not collected in the same order as the enumeration of the population census and proved impossible to link. Together, they represent about 17% of the decedents enumerated. Another 61,000 decedents in 1880 were added from physician reports and could also not be linked.

physicians, and attributed causes and classification systems reflected this lack of knowledge (Anderton and Leonard 2004).

Most importantly, the mortality censuses significantly undercounted deaths—perhaps 40% or more in some years—limiting their usefulness. Undercounts varied by age (deaths among infants and the elderly were underreported relative to deaths at other ages), across space (deaths in the South and West are believed to have been undercounted relative to deaths in other regions), and over time. The 1850 Census Superintendent James D. B. DeBow, for example, observed that the “varying ratios between the States, as drawn from the returns, show not so much in favor of or against the health of either as they do, in all probability, a more or less perfect report of the marshals [enumerators]. Thus it is impossible to believe Mississippi a healthier State [as indicated by the census returns] than Rhode Island, etc.” (U. S. Census Office 1855: 8). Two decades later, the introduction to the published mortality statistics in 1870 wryly noted that “if the value of the statistics of mortality in a census of the United States, taken under existing laws, depended upon the return of substantially the whole body of deaths occurring during the year covered by the enumeration, the results would not be worth the space occupied by publication, much less the expense of collection and compilation.”³

Motivated by a desire to improve the quality of the data, the 1870 Census Office began using death registration records from local and state boards of health where they existed. Although likely more complete, deaths recorded in the death registration records lack a family number, which facilitates linking decedents to their households of origin on the census population schedule. The Census Office in 1880 expanded this practice of using local death registration data and augmented the enumerator’s returns one other way. Early in the census year, the Census Office mailed death registration forms to

³ This text was also quoted in the 1880 introduction, which considered the sentiment to still apply. John S. Billings. “Report on the Mortality and Vital Statistics of the United States as Returned at the Tenth Census (June 1, 1880)”. p. xi.

known physicians on lists supplied by postmasters. Although only 37 percent of the forms were filled out and returned, the Census Office compared the returned physician's forms to enumerator schedules for relevant localities. If a physician's forms contained a decedent that was not on an enumerator's sheet, clerks appended the name and death information by hand. In some places the Census Office relied solely on health board records for the published tabulations in 1880 (this was the case for all of New Jersey, Massachusetts, the District of Columbia, and 19 large cities in other states).

In a useful description and evaluation of the aggregated returns from the mortality censuses reported by the Census Office, Gretchen Condran and Eileen Crimmins (1979) observed that many of the problems in the mortality censuses stemmed from the retrospective nature of the data collection. Deaths among individuals living in single-person households or the death of all members of a multi-person household resulted in no survivors to report the decedents' deaths to a census enumerator. Deaths of household heads that resulted in the dissolution of a household were also likely to have been under-reported. These omissions likely explained much of the under-reporting of deaths among older adults, who were more likely to live alone and more likely to be a head of household. Another consequence of the retrospective nature of the mortality census was recall error among respondents, who were less likely to report deaths occurring early in the previous year than deaths occurring in the months immediately prior to the census (Ferrie 1996). These findings are consistent with modern research on recall errors in survey research. As the length of time over which recall is required grows, people are more likely to collapse or telescope time (Kjellson, Clarke, and Gerdtham, 2014). A highly salient and rare event like mortality is unlikely to be forgotten entirely if the household survives, but recollection of the exact date in response to an enumerators' question is more likely.

One group of decedents—children aged 5 to 19—appears to have been relatively well reported. In the most highly cited research use of the mortality data collected by the census, Michael R. Haines relied on the rate of death among individuals aged 5-9, 10-14, and 15-19 to construct national life tables

for census years 1850-1900 (Haines 1979; 1998). Haines fitted the observed age-specific mortality rates to model life tables, which allowed him to extrapolate mortality rates in infancy and for older age groups. The method resulted in a series of life tables for the total and white populations of the United States by sex for each census year between 1850 and 1900. The results appear reasonable and do not suggest a serious underreporting of mortality among the age 5-19 reference population. According to Haines' "U.S." life tables, life expectancy for both sexes combined increased from 39.1 in 1850 to 41.7 in 1900, generally consistent with estimates from England and Wales and other comparable countries. Although covering only the years prior to each census, Haines' life tables represented a great improvement to previously available life tables and have been widely used by demographers and economists (Hacker 2010).

The value of the mortality censuses can be significantly enhanced using the individual-level data collected on the original returns. Census of mortality microdata, for example, can be used to select unique population subgroups for analysis, create custom tabulations to overcome limitations in the published statistics, and construct empirical models of mortality. Joseph Ferrie (1996) was the first researcher to analyze census of mortality microdata. Using a sample of 30 thousand decedents in the 1850 mortality census transcribed by the genealogical company Accelerated Indexing Systems, Ferrie constructed life tables for 10 subgroups of adult males (urban, rural, Northeast, Northwest, and South for both the native-born and foreign-born populations). The microdata allowed Ferrie to limit the number of deaths used in calculating age-specific death rates to those occurring within six months of the census (multiplied by two to yield an estimate of the annual number of deaths). In this way, he avoided relying on deaths in the period 6-12 months prior to the census, which suffered higher undercounts.

Mortality census microdata contain additional information—sex, race, occupation, cause of death, place of residence, etc.—that can be used to examine individual correlates of death. In a subsequent study, Ferrie (2003) relied on a larger sample of decedents in the 1850 and 1860 censuses of

mortality to evaluate the impact of occupation on male mortality in different age groups, controlling for nativity, migration, access to transportation, and region. Ferrie found little impact of occupation on all causes of mortality, but lower mortality from “consumption” (pulmonary tuberculosis) among laborers compared to white-collar workers and craftsmen, perhaps a result of their differing work environments.

The Manuscript Returns

The first census of mortality was one of six schedules completed by assistant marshals as part of the 1850 census. Schedule 1 was the census of the free population, schedule 2 was for the slave population, schedule 3 was for mortality, and the remaining schedules were for agriculture, industry, and social statistics. Presumably, enumerators visiting a household first completed the population census, then inquired whether the household had any members who died in the year prior to the census. If a death had occurred, the enumerator asked several additional questions about the decedent and entered the information on the mortality census. If there were no deaths, nothing was recorded in the mortality census.

The 1850 mortality census questions included the name, age, sex, color (race), whether free or slave, marital status, place of birth, occupation, cause of death, number of days ill, and the place, county and state of persons who died during the year preceding June 1, 1850. Additional questions were added in subsequent census years. Beginning with the 1870 mortality census, the sequential family number in the population census for the family reporting the death was recorded on the mortality census.

Figure 1 shows a partial example of a manuscript page for the 1880 mortality census. The form includes the family number on the population schedule of the reporting household; the name, age, sex, color (race), marital status, place of birth, place of father’s and mother’s birth, profession, month of death, disease or cause of death, how long a resident of the county, place the disease was contracted (if not the same as the place of death), and the name of attending physician for each decedent; and the

place, county, and state of the persons dying (recorded at the top of the form for each decedent on the page). In this example, the first decedent, Catherine Trimble, was aged 65 at death, and died in April from heart disease. The second, George Thompson, was aged 1 year and 3 months and died in November from spinal meningitis. The first seven decedents had family numbers less than 100, suggesting that less than one in ten families in the population census reported a death. One family—the Schroder family with family number 96 on the population census—reported two deaths, one for an unnamed female (“infant”) who died in July, aged 2 days—whose cause of death was listed as “infancy”—and Caroline Schroder, a female who died 3 months later aged 1 year and 3 months from diphtheria.

Manuscript returns for the mortality census were preserved by state historical agencies and research libraries. As noted earlier, however, not all have survived. In a significant minority of cases, the returns have been lost. Even when the manuscripts existed, access was more difficult than for the population schedules. The Census Bureau documented where the mortality schedules were believed to be located in a 2002 report on 210 years of American census taking (U.S. Census Bureau, 2002). Beginning with data donated by Family Search and Ancestry, we discovered that most of the mortality schedules that had been included on National Archives and Records Administration (NARA) microfilm reels were included in the database, as were most microfilmed records held by state archives, libraries, or historical societies.

Surprisingly, the Census Bureau report included a few errors about the existence of records. For example, the Florida mortality schedules are listed as appearing on NARA microfilm, along with other non-population schedules. When we obtained the listed microfilm, we discovered the mortality schedules had not been filmed. Conversely, Vermont’s mortality schedules were identified as existing in manuscript form only, for every year except 1870. Neither Florida nor Vermont was able to microfilm material on request. In both cases, the state agency did not have the capacity to undertake filming. Nor

was there an option to contract with private companies for microfilming, which may have been an option as recently as the early 2000s. Thus, we undertook two separate trips to each state to photograph the manuscripts, and provided free copies of the digital photographs to the state agencies who held the manuscripts, to assist with public access and preservation.

In other cases, *Measuring America* notes a NARA reel, but the data was significantly incomplete. We include this information in Table 1. We strove to locate all surviving manuscripts, and when necessary, paid for microfilming original schedules. For these records we conducted in-house data entry at the University of Minnesota. By this process we were able to include mortality data for Dakota Territory 1860-1880, Delaware 1860-1880, Florida 1850-1880, Indiana 1860-1880, Maryland 1870-1880, Mississippi 1870, Missouri 1860-1880, New Hampshire 1880, New Mexico 1860-1880, Oregon 1860-1880, Rhode Island 1860-1880, Texas 1860, Utah 1870-1880, Vermont 1860, West Virginia 1870-1880, Wisconsin 1880, and Wyoming 1870-1880. In contrast to much of the data obtained from Ancestry, these records are more likely to include complete information on occupation and other variables.

IPUMS full count datasets

The IPUMS full count datasets of the 1850-1880 mortality censuses were constructed from data provided by Ancestry, most of which were entered by data entry subcontractors in East Asia without oversight by IPUMS staff, supplemented with data entered by Minnesota Population Center employees for places we were able to locate additional surviving records. Our first tasks were to assemble these data, search for duplicates and missing records, make manual edits where needed, and then convert all string variables to numeric variables with IPUMS coding schemes (e.g., state, county, place, sex, race, marital status, occupation, etc.).

Table 2 shows the number of deaths and a selection of the more important variables available for analysis for each of the datasets. As shown in the last row of the table, the percentage of the

originally enumerated decedents included in the datasets ranges from a low of 85.2% in 1870 to a high of 98.3% in 1860. Researchers using datasets with a high percentage of missing records should be aware, therefore, that the number of deaths in an area may be low relative to the original enumeration, and that deaths documented in the database may be unrepresentative of all deaths. If the datasets are used to construct age-specific death rates, researchers should be careful to include the appropriate denominators of the at-risk population by excluding the population living in states and counties with no or incomplete surviving mortality data. Unfortunately, the mortality information published by the Census Office in the nineteenth century did not include cross-tabulations of deaths by county, so we are unable to confirm which counties in states with incomplete records have missing data. Obviously, large counties with no reported decedents have no surviving mortality census schedules, but counties with low death rates may have partial surviving information or simply benefitted from a relatively healthy environment in that particular year.

As discussed in an associated paper on the IPUMS full count 1850-1880 population census databases, the sheer number of unique string values overwhelmed our limited resources. Earlier IPUMS databases relied on human coding of every unique string value supplemented by imputation of the few remaining responses that were illegible or missing using “hot deck” procedures, in which missing values are replaced with an observed response from a similar individual. In contrast, the datasets for the mortality censuses have some responses that remain uncoded. Among the 161,508 unique string expressions for cause of death in the combined 1860-1880 census of mortality datasets, for example, an IPUMS research staff member manually coded 104,471 strings (64.7% of the unique cause of death strings). The remaining uncoded strings represent relatively few people, however. We focused our coding efforts on strings that appeared most frequently. Most of the uncoded responses were for strings unique to a single individual (e.g., only one individual had an occupation of “wastbruck on the head by a bliceman in attenpting to arret him scull brighted a hired w a from hours,” which remains uncoded).

Because the coded strings often apply to many individuals (e.g., 130,679 individuals had the cause of death string “consumption”), only 5.5% of decedents in the 1860-1880 mortality census datasets have a cause of death that was not classified. All datasets include the string variables, allowing users to develop strategies for coding or distributing these cases as they see fit.

We made no attempt to duplicate the unique occupational and cause of death classification systems used by the Census Office in the nineteenth century. Decedents’ occupations, when captured in the dataset, were coded using the 1950 Census Bureau’s occupational classification system—the most used IPUMS classification for historical research. Cause of death was coded using the International List of Causes of Death, revision 2 (1909). There were, of course, many challenges in coding cause of death, some of which arose from data entry errors and others from the poor understanding of diseases among nineteenth-century respondents. We recommend that users interested in analysis of cause of death become familiar with nineteenth-century nosology (for an example, see Anderton and Leonard 2004) and exercise caution interpreting the results.

As noted above, the Census Office was aware that the number of deaths reported in the mortality censuses was too low and took steps beginning with the 1870 census to supplement the returns with death registration data where it existed. One of the challenges of constructing the datasets was to identify the different types of mortality records (i.e., for some places we only have enumerator schedules, for some places we only have health board records, and for some places we have a mix of both). For a few states the datasets have more deaths than were tabulated by the census, suggesting that some deaths were recorded twice. We were able to remove some duplicates but struggled to identify and remove others. There are, for example, significantly more deaths in the 1860 dataset for Rhode Island and in the 1870 dataset for New York than was tabulated by the Census Office.

Despite the attempts by the 1870 and 1880 Census Office to supplement recorded deaths, both censuses underestimated mortality. In Figure 2, we compare the age-specific proportions dying (the q_x function in the life table) in Haines' (1998) life table estimates for the white population in 1870 to the 1870 mortality data. The mortality estimates for the IPUMS data are limited to the white population residing in the 33 states in 1870 whose aggregate number of deaths in the dataset was within plus or minus two percent of the number of deaths reported in published census returns. Unsurprisingly—since Haines relied on age 5-9, 10-14, and 15-19 death rates to fit the model, the results match closely in those age groups. In all other age groups, the mortality rates calculated with the census data are lower than those predicted by the model; significantly lower (40% or more) at age 0 and for age groups above 50. If Haines' model results are an accurate portrayal of true death rates, the overall number of deaths reported in the census were undercounted by 30.2%. Some of these missing deaths may have been in the earlier part of the year and thus were more prone to be forgotten by respondents. Because we have microdata, it is easy to restrict the analysis to deaths occurring in the previous six months, which were presumably more likely to be remembered. The exercise supports the hypothesis that deaths earlier in the year are more likely to be underreported, but the adjusted proportions dying for each age group is only marginally higher; overall, the adjusted number of overall number of deaths is 27.8% below the deaths suggested by the model. Interestingly, relying solely on deaths in the previous six months results in a lower death rate for infants, likely because the six months prior to the census (December 1869 – May 1870) excludes summer months, when infant mortality rates are typically higher (see discussion in Oris et al. 2023).

Linked Datasets, 1870 and 1880

The research potential for the IPUMS mortality datasets is increased by linking decedents to their households of origin in the IPUMS full count datasets of the population censuses (Ruggles et al. 2021). Occupations of fathers, for example, can be included in analyses of child deaths. Wealth data—

uniquely collected in the 1850 (real estate wealth only) and 1860-1870 censuses (real estate and personal estate)—is also available, along with other household measures (children in school, mothers in the paid labor force, etc.).

Until recently, linking the mortality and population censuses was an arduous task. In the 1850 Census Compendium, Census Superintendent J. D. B. De Bow complained about physical separation of the mortality schedules from the population census, which made linking difficult: “nor can the deaths of individuals be associated with families, and with the remainder living in families, without almost impracticable labor” (DeBow 1851, p. 14). Linking is now feasible, however, thanks to the way the census was conducted, the creation of full count datasets of the population censuses, and new automatic linking methods (e.g., Helgertz et al. 2022).

The manner in which the census was conducted in the nineteenth century is one reason linking the population census with the mortality census is a possibility. Census enumerators physically visited each household, going door-to-door on foot or by horseback and completing each relevant census schedule in turn. In some households—those with deaths, slaves, and agricultural products—enumerators completed four separate census schedules on the same visit. In others, they completed fewer. Although every family was enumerated in the population census, and numbered sequentially in their order of visitation, only families with deaths, slaves, and farm products were enumerated in the mortality, slave, and agriculture censuses. Although addresses were not recorded until the 1880 census, and then only for a minority of households in urban areas, the sequential order of the enumerator’s visitation is preserved in all four schedules. In most cases, a decedent enumerated after another decedent on the mortality census was reported by a family on the population census that was enumerated after the first decedent’s family, although there are often families in between. By the absence of a death on the mortality censuses, it can be inferred that families in between the two families in the population census that reported a death had no deaths to report. Theoretically,

therefore, individuals' surnames and their order of enumeration should allow decedents in the IPUMS complete-count mortality datasets to be linked to the IPUMS complete-counts population datasets, at least for decedents whose surname was shared by a member of the reporting household.

In practice, we found the task of linking the 1850 and 1860 mortality records to the 1850 and 1860 IPUMS complete-count population databases too difficult when conducted on a large scale with automatic linking methods. Although most mortality and population records appeared to be in roughly the same sequential order in the data, we encountered considerable ambiguity in the population data. One difficulty was the large number of households in the nineteenth century in close proximity to other households with the same surname (likely relatives)(Nelson 2020). Different spellings of the same surname in the different datasets also presented problems determining the correct link. In some of the ambiguous situations the link might be obvious (e.g., an elderly male decedent, if currently married at the time of death, should link to a household with an elderly woman of approximately the same age who does not have an obvious co-resident spouse). But any links would be relatively unrepresentative and would involve significant guesswork. Such links are perhaps feasible when conducted by an individual researcher on a case-by-case basis, but difficult to automate and too expensive to conduct at scale. Without common family numbers in the mortality and population schedules, we judged the small number of links we were able to make far too tenuous to construct a reliable dataset.

Beginning with the 1870 census of mortality, enumerators recorded the decedent's family number from the population schedule, providing a direct link and theoretically removing the need to link using surnames and sequential ordering. Even here, however, we found the task of making automatic links difficult. Data entry errors of the family number in either the population or mortality datasets were not uncommon. Typically, family numbers were not unique within a given county in the population data. Different enumerators in the same county, of course, started with family sequence number one. It was also common for enumerators to restart the sequential numbering of families when

moving to a different area of a county (a different township, beat, district, municipality, etc.) In 1880, the Census Office defined enumeration districts (EDs), which were separately numbered. Within each ED, family numbers were to be unique for each family. Thus in 1880, we can match by state, county, ED, and family number. For 1870, which did not record an enumeration district number, we developed a method of inferring enumeration districts. We describe the 1880 procedures first.

For 1880, in addition to matching decedents to their households of origin using state, county ED, and family number, we also matched on surname similarity. Our initial potential links file consisted of matches between the decedent's surname and any household in the same state and county with a similar surname (0.9 Jaro-Winkler threshold). After some experimentation, we decided to only accept links if they matched for all four characteristics (state-county-ED-family number) and met the 0.9 Jaro-Winkler surname similarity threshold. Although matches with just state-county-ED-family number were feasible and worked in most cases, we judged the number of false positives to be too high. Requiring a Jaro-Winkler surname similarity match in addition to EDs and family number matches resulted in a low linkage rate—just 34.4% of all decedents were matched to their households of origin in the 1880 population census—but high accuracy among the matches that were made.

Many of the decedents we failed to link in 1880 had missing family numbers or enumeration districts. This was true for all decedents obtained from Health Board records and the decedents added from the physician's reports. Examination of failed links among the decedents with both family numbers and enumeration districts revealed frequent transcription errors. The population and mortality records were transcribed by different data entry operators on different dates, increasing the probability of data entry errors in family numbers, enumeration district numbers, and surnames. We also failed to find the true link in cases where there was no one in the population household that had the same surname as the decedent within the required Jaro-Winkler threshold.

Although the 1870 mortality census included fewer records entered from local vital registration systems—and therefore has a higher percentage of cases with family numbers—an added difficulty is the absence of ED numbers. We dealt with this limitation by constructing pseudo enumeration districts in the 1870 population data. The basic premise was that a single manuscript page of mortality data (a maximum of 35 lines) would correspond to a single constructed ED in the 1870 population data. We established clusters of potential links between specific mortality pages and constructed EDs, and ultimately accepted links that matched on state, county, family number, and constructed enumeration district (in addition to the 0.9 Jaro-Winkler threshold for surnames). This resulted in a 52.1% linkage rate. The linkage rate was marginally higher at younger ages, averaging 57% for ages under 15, 49% for ages 15-59, and 46% for older ages.

The identification of decedents in the linked dataset, of course, is a joint product of whether their deaths were reported by a family member and whether the decedents could be linked to a family. In Figure 2 above we also show the estimated life table proportion dying for the white population in the linked dataset. As can be seen in the figure, the age-pattern of mortality is similar to the assumed standard and to rates estimated from the non-linked dataset, but lower. The proportion of the population dying in each age group was approximately 50% lower than Haines' life table standard for age groups between age 1 and 25, and even lower for infants (an estimated 71% below the proportion suggested by the standard).

Modeling child mortality

As an example of the kind of analysis possible with the linked dataset, we constructed a logistic regression model predicting the deaths of children aged 0-14 in the 1870 linked mortality data, and separate models for children aged 0, 1-4, 5-9, and 10-14. Our objective was to determine whether a significant mortality gradient existed prior to the long-term decline in mortality that began in the United

States circa 1880. During this period parents possessed limited knowledge about the causes of diseases and preventative measures, public health agencies—in the few places in which they existed—were small and largely ineffective, and physicians had limited means to treat sick patients. In this environment wealthy parents may not have been better able to protect their children from the risk of death relative to poorer parents from the highly communicable diseases that were the leading killers. The social science literature is mixed on the existence and size of a gradient prior to the mortality transition. Although some researchers assert that differences in economic wellbeing and socioeconomic status are “fundamental causes” of mortality differentials (e.g., Link and Phelan 1995; Antonovsky 1967), research in Sweden and other places (e.g., Jaadla et al. 2020; Dribe and Karlsson 2022) suggests that a mortality gradient was small and inconsistent or did not emerge until the twentieth century, when new knowledge about the causes and prevention of diseases became more widespread and the practice of medicine more effective, allowing parents with greater education and financial means to leverage those resources on behalf of their children. As an example of the modest impact that education and socioeconomic status had on child mortality in the past, Condran and Preston noted that children of physicians in the late nineteenth-century United States enjoyed only a 6 percent lower risk of death than children of other fathers with other occupations, all else being equal, while children of teachers had no advantage (Condran and Preston 1994). With the direct measurement of parental wealth in 1870, the linked census of mortality dataset represents an ideal source to examine this question.

We limited the analytical sample to white and black children living in households headed by males with spouses present who were aged 15-59 in the population census (in most cases, the spouses will be the mothers of the children in the model). We combined the household head’s real and personal estate wealth and grouped total wealth into three categories: no wealth, moderate wealth (\$100-\$2999), and high wealth (\$3000 and over). Just over 30% of household heads reported no real or personal wealth, 51% reported moderate wealth, and the remaining 19% reported high wealth. Because

mothers' and fathers' birthplaces and literacies were strongly intercorrelated, we created couple-level variables for nativity (both spouses native-born of native-born parents, both spouses foreign-born, and spouses of mixed native-born and foreign-born parentage) and literacy (both spouses literate, or one or both spouses illiterate), and whether the household head was a farmer. We created four dummy variables for size of place, ranging from rural areas and places with less than 2,500 residents, small cities with 2,500-24,999 residents, medium-sized cities with 25,000-99,999 residents, and large cities with 100,000 or more residents. We also created dummy variables for each census region and included them in the models.

The results are shown in Table 3 for the 51,474 decedents linked to the population dataset. The exponentiated coefficients for each independent variable show the likelihood of a child death relative to the reference group (1.0). Looking first at model 1 for all children aged 0-14, we see that Black children and children of foreign-born parents suffered higher mortality than white parents of native-born parentage, while children of literate parents and children in farm families experienced lower mortality than children of illiterate parents and those living in households headed by a non-farmer. Place of residence clearly mattered. All else being equal, children living in urban areas experienced a much higher risk of death, with children in small and medium-sized cities experiencing a 42% and 50% higher risk of death relative to children living in rural areas, respectively, and children in large cities of 100,000 or more residents experiencing an 88% higher risk of death.

The wealth dummy variables indicate the presence of a significant gradient in child mortality by household wealth. Compared to children in households whose head reported no real or personal estate wealth, children in households with high wealth experienced about a 10% lower risk of death. This result agrees with a recent analysis by J. David Hacker, Martin Dribe, and Jonas Helgertz (2023), which used census panel data in the IPUMS Multigenerational Longitudinal Project for the 1850-1860, 1860-1870, and 1870-1880 intercensal intervals to examine the association between parental wealth and child

mortality. They reported a smooth mortality gradient by wealth decile, with children of parents in the top decile experiencing about 15-20% lower child mortality rates relative to couples with no wealth, all else being equal.

Turning to models for different age groups of children, we see that while the correlations are similar for different variables at most ages, there are several interesting differences. The model results suggest that infants living in households with heads in the highest wealth group suffered slightly higher mortality than the reference group of households headed by men with no reported wealth, all else being equal. Although there could be some explanation for these unexpected results—perhaps wealthy households were more likely to rely on wet nurses and therefore experience higher child mortality, for example—we suspect that the results for children aged 0 to be unreliable due to biases in reporting. As noted above, the number of infant deaths linked to the population is approximately 71% below the number suggested by Haines’ 1870 life table. If under-reporting of infant deaths varied by household wealth, the results could be spurious. Likewise, nineteenth-century observers such as De Bow were clearly surprised that the results indicated much higher mortality rates in the Northeast than in the South. The significantly lower risk of mortality in all regions relative to the Northeast could reflect more complete reporting of mortality in the Northeast rather than higher rates of child mortality. We urge researchers to be cautious in their analyses and consider using area fixed effects models to control for area differences in coverage and other unobserved heterogeneities.

Conclusion

New nineteenth-century population, mortality, and slave full count datasets from IPUMS are important new resources to study social and demographic change in the United States. This article describes four new mortality datasets for census years 1850, 1860, 1870, and 1880 for the study of mortality in the late nineteenth century, including examination of its levels, causes, and correlates.

Combined, these datasets contain information on over 1.8 million decedents. The 1870 and 1880 datasets include links to the population censuses, supporting additional analyses. Because so little individual-level data on deaths are available for the United States, these datasets represent a major new foundation for the study of mortality prior to the onset of the mortality transition, which commenced in the United States circa 1880 (Hacker 2010).

With the notable exception of states and parts of states whose mortality censuses did not survive, all four of the mortality datasets are “full count,” facilitating the study of subpopulations and small areas. These data are not without their problems, however, and users should be cautious about undercounts and potential biases. The number of deaths reported are clearly too low, likely the result of the retrospective nature of the censuses. Unfortunately, there are no obvious sources for assessing whether under-reporting varied over time, place, and population group. With care, however, we expect that these new datasets will be leveraged to better understanding of population dynamics in an era poorly documented by quantitative data.

Acknowledgments

This research was supported in part by funding from the Minnesota Population Center (P2C HD041023) and by grants from the Eunice Kennedy Shriver National Institute for Child Health and Human Development (R01-HD060676-01 and R01-HD082120-01). Many individuals at the Minnesota Population Center contributed to the project, but we wish to especially acknowledge the work of Ronald Goeken, who contributed a significant amount of effort to all aspects of the project over a five-year period. Bigyan Khanal provided some initial programming. Some of the information in this paper was originally presented as Ronald Goeken, J. David Hacker, Lap Huynh, and Bigyan Khanal, “An Inquiry into Some Points of Record Linkage: Linking 19th Century U.S. Decennial Population Records to the Mortality and Slave Schedules,” presented at “Putting the Pieces Together: Promise, Programs and Pitfalls in Linking Historical and Contemporary Records,” at the Center for Economic History, Northwestern University, May 17-19, 2019.

References

- Anderson, Margo J. 1988. *The American Census: A Social History* (New Haven: Yale University Press).
- Anderton, Douglas L and Leonard, Susan Hautaniemi. 2004. Grammars of Death: An Analysis of Nineteenth-Century Literal Causes of Death from the Age of Miasmas to Germ Theory. *Social Science History* 28:1, 111-143.
- Antonovsky, Aaron (1967) "Social class, life expectancy and overall mortality." *Milbank Memorial Fund Quarterly* 45 (2): 31–73.
- Condran, G. A. and E. Crimmins. 1979. A description and evaluation of mortality data in the federal census: 1850-1900. *Historical Methods* 12: 1-23.
- Condran, Gretchen A., and Samuel H. Preston (1994) "Child mortality differences, personal health care practices, and medical technology: The United States, 1900–1930," in Lincoln C. Chen, Kleinman Arthur, and Norma C. Ware (eds.) *Health and Social Change in International Perspective*. Cambridge, MA: Harvard University Press: 171–224.
- J.D.B. Debow. 1853. *The Seventh Census of the United States: 1850*. Washington, D.C.: *Robert Armstrong, Public Printer*.
- Dribe, Martin and Omar Karlsson (2022) "Inequality in early life: Social class differences in childhood mortality in southern Sweden, 1815–1967." *Economic History Review* 75 (2): 475–502.
- Ferrie, J. P. 1996. A new sample of males linked from the public use microdata sample of the 1850 U.S. federal census of population to the 1860 U.S. federal census manuscript schedules. *Historical Methods* 29: 141-156.
- Ferrie, J. P. 2003. The rich and the dead: Socioeconomic status and mortality in the United States, 1850-1860. In *Health and Labor Force Participation over the Life Cycle: Evidence from the Past*, edited by D. L. Costa, 11-50. Chicago, IL: The University of Chicago Press.
- Hacker, J. David. 2010. Decennial Life Tables for the White Population of the United States, 1790-1900. *Historical Methods* 43:3, 45-79.
- Hacker, J. David, Dribe, Martin, and Helgertz, Jonas. 2023. Wealth and child mortality in the nineteenth-century United States: Evidence from three panels of American couples, 1850-1880. *Social Science History* 47:3, 333-366.
- 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. 1998. Estimated life tables for the United States, 1850-1910. *Historical Methods* 31: 149-169.

Helgertz, Jonas, Price, Joseph, Wellington, Jacob, Thompson, Kelly J., Ruggles, Steven, and Fitch, Catherine A. 2022. A new strategy for linking U.S. historical censuses: A case study for the IPUMS multigenerational longitudinal panel. *Historical Methods: A Journal of Quantitative and Interdisciplinary History* 55 (1): 11–28.

Helgertz, Jonas, Ruggles, Steven, Warren, J. Robert, Fitch, Catherine A., Goeken, Ronald, Hacker, J. David, and Sobek, Matthew. 2020. IPUMS Multigenerational Longitudinal Panel: Version 1.0 [Data set]. Minneapolis: IPUMS.

Jaadla, Hannaliis, Ellen Potter, Sebastian Keibek, and Romola Davenport (2020) “Infant and child mortality by socio-economic status in early nineteenth-century England.” *Economic History Review* 73 (4): 991–1022.

Kjellsson, Gustav, Philip Clarke, and Ulf-G Gerdtham. 2014. "Forgetting to remember or remembering to forget: a study of the recall period length in health care survey questions." *Journal of Health Economics* 35:34-46.

Link, Bruce G., and Jo Phelan (1995) “Social conditions as fundamental causes of disease.” *Journal of Health and Social Behavior.* 35 (extra issue): 80–94.

Nelson, Matt A. 2020. The decline of patrilineal kin propinquity in the United States, 1790–1940, *Demographic Research* 43:18, 501-532

Preston, S. H., and M. R. Haines. 1991. *Fatal years: Child mortality in late nineteenth-century America*. Princeton, NJ: Princeton University Press.

Ruggles, Steven, Fitch, Catherine A., Goeken, Ronald, Hacker, J. David, Nelson, Matt A., Roberts, Evan, Roberts, Schouweiler, Megan, and Sobek, Matthew. 2021. IPUMS Ancestry Full Count Data: Version 3.0 [dataset]. Minneapolis, MN: IPUMS. <https://doi.org/10.18128/D014.V3.0>

Smith, Daniel S. (1983) “Differential mortality in the United States before 1900.” *Journal of Interdisciplinary History* 13 (4): 735–59.

United States Census Bureau. 2002. *Measuring America: The Decennial Censuses from 1790 to 2000*. Washington, DC.

Vinovskis, M. A. 1972. Mortality rates and trends in Massachusetts before 1860. *Journal of Economic History* 32: 184–213.

Vinovskis, M. A. 1978. The Jacobson life table of 1850: A critical re-examination from a Massachusetts perspective. *Journal of Interdisciplinary History* 8: 703-724.

Table 1. Location and survival of census mortality schedules

This table is adapted from the information provided by the Census Bureau in *Measuring America* (2002). Where applicable, we note important differences from the information provided in *Measuring America*. This listing provides, by state and year, the available mortality schedules. A cell that is blank indicates that there was no mortality data collected in that year in that area. Where the schedule has a National Archives publication number (M, T, GR, A, etc.) that number is listed. We note where we discovered significant shortfalls from the published mortality totals. If the publication was issued by a state archives or other organization, that organization is listed as the originator. Where there is no microfilm publication and the mortality schedule is available in book form only, that is indicated in the individual entry. If “manuscript” is indicated, the schedule has not been published and is available only at the holding institution.

State	1850	1860	1870	1880	1885
Alabama	Alabama Dept. Of Archives and History (ADAH)	ADAH	ADAH	ADAH	
Arizona		New Mexico State Records Center and Archives (NMSRCA)	T655	T655	
Arkansas	Arkansas History Commission (AHC)	AHC	AHC	AHC	
California	UC Berkeley Bancroft Library (BL)	BL	BL Partially complete	BL	
Colorado			T655	T655	M158
Connecticut	Connecticut State Library (CSL)	CSL	CSL	CSL	
Delaware	A1155	A1155	A1155	A1155	
District of Columbia	T655	T655	T655	T655	
Florida	T1168 Not on microfilm. Manuscript only.	T1168 Not on microfilm. Manuscript only.	T1168 Not on microfilm. Manuscript only.	T1168 Not on microfilm. Manuscript only.	M845
Georgia	T655	T655	T655	T655	

State	1850	1860	1870	1880	1885
Idaho	(book form)		Idaho State Historical Society		
Illinois	T1133	T1133	T1133 54% of records missing	T1133	
Indiana	Indiana State Library (ISL)	ISL Digitized and online	ISL Digitized and online	ISL Digitized and online	
Iowa	A1156	A1156	A1156	A1156	
Kansas		T1130	T1130	T1130	
Kentucky	T655 74% of records missing	T655	T655	T655	
Louisiana	T655	T655	T655	T655	
Maine	Maine State Archives (MSA)	MSA	MSA	MSA	
Maryland	Maryland State Law Library (MSLL)	MSLL Digital images available	MSLL Digital images available	MSLL Digital images available	
Massachusetts	GR19 T1204	GR19 T1204	GR19 T1204	T1204	
Michigan	T1163	T1163	T1163	T1163	
Minnesota	Minnesota Historical Society (MHS) (manuscript)	MHS	MHS	MHS	
Mississippi	Mississippi Dept. Of Archives and History (MDAH)	MDAH	MDAH	MDAH 50% missing	

State	1850	1860	1870	1880	1885
Missouri	State Historical Society of Missouri (SHSM)	SHSM Microfilm available from FamilySearch	SHSM Microfilm available from FamilySearch. Partially complete	SHSM Microfilm available from FamilySearch	
Montana			GR6	GR6	
Nebraska		T1128	T1128	T1128	M352
Nevada			Nevada Historical Society (NHS) (manuscript)	NHS (manuscript)	
New Hampshire	New Hampshire State Library (NHSL)	NHSL Microfilm loanable	NHSL Microfilm loanable	NHSL Microfilm loanable	
New Jersey	GR21	GR21	GR21	GR21	
New Mexico	NMSRCA	NMSRCA Digital images available	NMSRCA Digital images available	NMSRCA Digital images available	M846
New York	New York State Archives (NYSA)	NYSA	NYSA	NYSA	
North Carolina	GR1	GR1	GR1	GR1	
North Dakota		South Dakota State Historical Society (SDSHS) Microfilm cannot be loaned or copied	SDSHS Microfilm cannot be loaned or copied	SDSHS Microfilm cannot be loaned or copied	State Historical Society of North Dakota (manuscript)
Ohio	T1159 33% of records missing	T1159	T1159 Data appears to be entirely missing	T1159 Data only survives for some counties	

State	1850	1860	1870	1880	1885
Oregon	Oregon State Library (OSL)	OSL Microfilms are loanable	OSL Microfilms are loanable	OSL Microfilms are loanable	
Pennsylvania	T956	T956	T956 25% missing	T956	
Rhode Island	Missing data	Missing data	Rhode Island State Archives Digital images available	RISA Digital images available	
South Carolina	GR22	GR22	GR22	GR22	
South Dakota		SDSHS	SDSHS	SDSHS	GR27
Tennessee	T655	T655	Missing data	T655	
Texas	T1134 28% of records missing	T1134	T1134 GR7	T1134	
Utah	(book form)	(book form)	GR7 State	Missing data	
Vermont	Vermont Dept. of Libraries (VDL) (manuscript)	VDL (manuscript)	GR7	VDL (manuscript)	
Virginia	T1132 Only partially complete	T1132	T1132	T1132	
Washington	OSL	A1154	A1154	A1154	
West Virginia	West Virginia Dept. Of Archives and History (WVDAH)	WVDAH	WVDAH	WVDAH	

State	1850	1860	1870	1880	1885
Wisconsin	State Historical Society of Wisconsin (SHSW)	SHSW	SHSW	SHSW	
Wyoming			(book form)	(book form)	

Table 2. Partial list of Variables in the IPUMS Mortality Census Datasets, 1850-1880

Census year	1850	1860	1870	1880
		Restricted use only	Restricted use only	Restricted use only
Name of deceased person				
Color	X	X	X	X
Sex	X	X	X	X
Age	X	X	X	X
Free or slave	X	X		
Marital Status	X	X	X	X
Married or widowed	X	X	X	
Single, married, widowed or divorced				X
Place of birth	X	X	X	X
Parentage (mother, father of foreign birth)			X	X
Occupation	inc.	inc.	inc.	inc.
Month of death	X	X	X	X
Cause of death	X	X	X	X
Number of days ill	X	X		
Length of time resident in country				X
Name of place disease was contracted				X
Name of attending physician				X
State	X	X	X	X
County	X	X	X	X
Place (Township, district, city, etc.)	X	X	X	X
Family number			inc.	inc.
Historical ID of reporting household in population census			X	X
Number of decedents in dataset	298,870	387,206	419,927	728,858
Number of decedents in published data	323,098	394,153	492,263	756,898
Percentage of decedents in microdata	92.5%	98.2%	85.3%	96.3%
Number of decedents linked to family in population census			209,557	251,037
Percentage of decedents linked to population census			49.9%	34.4%

Notes: Restricted versions of the data are available for each nineteenth century census. The restricted versions include names, street address when available, and the input data (strings and coded). Accessing these data does require specific stipulations in order to use, and interested users should contact ipums@umn.edu or ipumsres@umn.edu to request access to these data. Variables denoted by "X" are census questions with available data in a given year and coded by IPUMS. Variables denoted by "C" were constructed using logical rules. "Inc." indicates census questions significantly incomplete or with significant errors in the dataset. Some of the original manuscript returns do not survive or could not be located and processed. States with a significant amount of missing mortality data include California (1870 [partial]); Illinois (1870 [partial]); Kentucky (1850 [partial]); Mississippi (1880 [partial]); Missouri (1870 [partial]); Nebraska (1870 [partial]); New Mexico (1850, 1860, 1870, and 1880); Ohio (1850 [partial], 1870, 1880 [partial]); Pennsylvania (1870 [partial]); and West Virginia (1860). Users should exercise caution when combining mortality census data with population census data to construct mortality rates.

Table 3. Logistic Regression death of child in household, 1870 linked mortality dataset (odds ratios)

Model number	(1)	(2)	(3)	(4)	(5)
Age of children	0-14	0	1-4	5-9	10-14
<u>Race of Child</u>					
<i>Race white</i>	<i>ref.</i>	<i>ref.</i>	<i>ref.</i>	<i>ref.</i>	<i>ref.</i>
<i>Race Black</i>	1.075 ***	1.127 ***	1.149 ***	1.134 **	1.352 ***
<u>Characteristics of Household Head</u>					
Occupation					
<i>Non-farmer</i>	<i>ref.</i>	<i>ref.</i>	<i>ref.</i>	<i>ref.</i>	<i>ref.</i>
<i>Farmer</i>	0.748 ***	0.794 ***	0.712 ***	0.750 ***	0.802 ***
Total Wealth (combined real and personal)					
<i>No wealth</i>	<i>ref.</i>	<i>ref.</i>	<i>ref.</i>	<i>ref.</i>	<i>ref.</i>
<i>Moderate wealth (\$100-\$2999)</i>	0.970 **	1.019	0.990	0.980	1.005
<i>High wealth (\$3000 and up)</i>	0.904 ***	1.064 *	0.976	0.947	0.882 *
<u>Characteristics of Household Head & Spouse</u>					
Nativity					
<i>Both native-born of native-born parents</i>	<i>ref.</i>	<i>ref.</i>	<i>ref.</i>	<i>ref.</i>	<i>ref.</i>
<i>Mixed or second generation</i>	1.090 ***	1.056 *	1.069 **	0.992	0.966
<i>Both foreign born</i>	1.096 ***	1.001	1.078 ***	1.027	0.938
Literacy					
<i>One or both Illiterate</i>	<i>ref.</i>	<i>ref.</i>	<i>ref.</i>	<i>ref.</i>	<i>ref.</i>
<i>Literate (both can read and write)</i>	0.888 ***	0.883 ***	0.898 ***	0.883 ***	0.898 *
<u>Characteristics of place of residence</u>					
Rural-Urban					
<i>Rural areas/places with less than 2500 pop</i>	<i>ref.</i>	<i>ref.</i>	<i>ref.</i>	<i>ref.</i>	<i>ref.</i>
<i>Places with 24,999 pop.</i>	1.418 ***	1.450 ***	1.412 ***	1.334 ***	1.278 ***
<i>Places with 25,000-99,999 pop.</i>	1.499 ***	1.460 ***	1.499 ***	1.456 ***	1.036
<i>Places with 100,000+ pop.</i>	1.875 ***	1.773 ***	1.972 ***	1.738 ***	1.257 **
Census Region					
<i>Northeast</i>	<i>ref.</i>	<i>ref.</i>	<i>ref.</i>	<i>ref.</i>	<i>ref.</i>
<i>Midwest</i>	0.918 ***	0.915 ***	0.940 **	0.930	0.868 **
<i>South</i>	0.700 ***	0.722 ***	0.697 ***	0.716 ***	0.682 ***
<i>West</i>	0.452 ***	0.328 ***	0.473 ***	0.919	0.690
Number of at-risk children	51,474	21,885	20,283	5,828	3,478
Number of observed deaths	6,952,501	557,486	2,125,615	2,190,115	2,079,285

Notes: *** p<0.001, ** p<0.010, * p<0.050. Universe includes all male-headed households with spouses aged 15-59 present in the household. All models include controls for spouse's age.

Figure 1. Partial image of an original manuscript page for the 1880 census of mortality

[7-222.]

Page No. 1
 Supervisor's Dist: No. 3
 Enumeration Dist: No. 37

Note A.—The Census Year begins June 1, 1879, and ends May 31, 1880.
 Note B.—In making entries in columns 6, 7, and 8, an affirmative mark only will be used, thus /, except in the case of Divorced persons, column 8, when the letter "D" is to be used.
 Note C.—For instructions relative to the entries in column 14, see back of this Schedule.
 Note D.—In column 17, note distinctly if no Physician was in attendance, thus (None.)

SCHEDULE 5.—Persons who DIED during the Year ending May 31, 1880, enumerated by me in Hamilton, in the County of Butler, State of Ohio, Lucius B. Potter, Enumerator.

Name of the person deceased.	Personal Description.			What was the civil condition of the person when he died?		NATIVITY.			Profession, Occupation or Trade. (Not to be asked in respect to persons under 10 years of age.)	The month in which the person died.	Disease or cause of death.	If the disease was not contracted prior to death, state the place.	Name of attending Physician.			
	Age at last birthday. If under 1 year, state number of months. If under 1 month, give date in fraction.	Sex—Male (M), Female (F).	Color—White (W), Black (B), Indian (I).	Single /	Married /	Divorced D.	Widowed /	Place of birth of this person, naming the State or Territory of the U. S., or the country, if of foreign birth.						Where was the Father of this person born? (As in column 9.)	Where was the Mother of this person born? (As in column 9.)	
1.	2.	3.	4.	5.	6.	7.	8.	9.	10.	11.	12.	13.	14.	15.	16.	17.
30 Trimble Catherine	65	F	W	/				Ireland	Ireland	Ireland	3 April	Heart Disease	40	Falloon		
51 Thompson Geo C.	1 1/2	M	W	/				Ohio	Ohio	Ohio	24 Nov	Spinal Meningitis	17 1/2	Carley		
62 Riet Carrie	2	F	W	/				Ohio	Germany	Ohio	2 June	Scarlet Fever	2	Hutton		
78 Schwart Christiana	3	F	W	/				Ohio	Germany	Ohio	16 Aug	Scarlet Fever	3	Hansen		
93 Rupp Amanda	1 1/2	F	W	/				Ohio	Ohio	Ohio	16 July	Scarlet Fever	1 1/2	Hellman		
96 Schroeder Infant	4 days	F	W	/				Ohio	Germany	Ky	7 July	Infancy	0	None		

Figure 2. Proportion of population dying in age interval, white population in 33 selected states and territories with complete mortality census data, both sexes combined

