Health, gender and mobility: Intergenerational correlations in longevity over time

John Parman*

March 27, 2016

Abstract

Changes in intergenerational mobility over time have been the focus of extensive research. However, existing studies have been limited to studying males and intergenerational correlations in outcome variables that often lack clear welfare implications. This paper introduces a new approach to estimating intergenerational mobility that relies on health measures rather than occupational measures to assess the strength of the relationship between the outcomes of parents and their children. Health measures provide a metric for intergenerational mobility that can be consistently interpreted over time and across genders. Using a new intergenerational dataset constructed by linking individuals' death certificates to those of their parents, I find that a son's life span is strongly correlated with his father's and that this correlation has strengthened over time. Daughter's life span shows a similarly strong relationship with mother's life span that has remained relatively stable over the past century. Differences in life span are shown to correlate with occupational status and occupational transitions from one generation to the next.

1 Introduction

A central question about the development of the American economy has been how the economic gains experienced by the economy as a whole were distributed across the population. Whether the American dream, the ability to rise from humble origins to riches, was a reality and whether it still exists today has been a question generating research across a wide range of disciplines. Attempts to quantify the extent of American intergenerational mobility have produced provocative but ultimately limited results both because of data limitations and the difficulties in interpreting the measures of mobility than can be estimated.

*jmparman@wm.edu, Department of Economics, College of William & Mary and NBER; This paper has greatly benefited from discussions with Trevon Logan and the results of our joint work on race and mortality utilizing similar death records. I thank the Economic History Association for supporting the data collection through a Cole Grant and the Schroeder Center for Health Policy for providing resources for research assistants. I am grateful for comments from seminar participants at the University of Michigan and Gettysburg College and at the Southern Economic Association annual meetings. This is a very, very preliminary draft, please do not cite without author's permission. For the most recent version of the paper, please see http://jmparman.people.wm.edu.

This paper takes a new approach to measuring intergenerational mobility, one that circumvents several of the limitations of prior mobility measures. I measure mobility by the strength of the relationship between the long term health outcomes of children and their parents, specifically by estimating the correlation between a child's life span and that of her parents. This approach addresses two major constraints faced by existing measures of historical mobility focused on occupational transitions. First, occupational transitions often have ambiguous welfare implications. A longer life span has clear welfare implications compared to transitions between broad occupational categories. Life span also has a consistent interpretation over time and across genders, a longer life means the same thing for a male or female in the nineteenth century or the twentieth century. Occupations have no such clear interpretation. A woman with no occupation does not necessarily have the same socioeconomic status as a man with no occupation. A nineteenth-century farmer may have an entirely different job and status than his twentieth-century counterpart.

The second advantage of using health to measure mobility is that it enlarges the scope of historical mobility estimates to encompass females. Occupational mobility studies and, to a lesser extent, wealth mobility studies focus out of necessity on sons and their fathers. The techniques to link children to their parents and the variables used to measure mobility have been inapplicable to historical female populations. Focusing on life span not only uses a measure that is equally meaningful for males and females, it also relies on death certificates which report maiden names making it possible to match adult females to their parents. This is not possible with other historical data sources. Given the dramatic differences in the evolution of male and female labor market and health outcomes over the past two centuries, the ability of health measures to provide mobility estimates for both males and females is a major contribution to our understanding of historical mobility patterns.

These newly constructed intergenerational health data present evidence that is largely complementary to existing measures of historical mobility rates. I find strong correlations between an individual's longevity and that of his or her parents, both for males and females. This correlation has been getting stronger for males over the past century but has remained relatively stable for females over time. The decrease in mobility for males is consistent with the existing literature on declining occupational and income mobility over time in the United States. Longevity is correlated with occupational status and changes in longevity across generations are correlated with changes in occupational status across generations with transitioning out of an unskilled occupation a strong predictor of an increase in longevity relative to one's parents. These results suggest that longevity and other long term health outcomes offer a promising new direction for the estimation of historical mobility patterns.

2 Approaches to Estimating Historical Mobility Rates

Modern intergenerational mobility can be captured by a variety of convincing measures as a result of the creation of large longitudinal datasets in several developed countries including the United States. These data allow for comparing the incomes and occupations of sons and daughters to their parents. A standard approach of these studies is to focus on estimating intergenerational income elasticities, with a higher intergenerational income elasticity implying a lower level of intergenerational mobility.[1] The results typically estimate an elasticity of between 0.3 and 0.4 between the earnings of sons and their fathers (see Solon (1999) for a review of several of these studies). Cross-country comparisons of these elasticities suggest that the United States displays similar or lower mobility rates than other developed countries (Solon, 2002; Aaberge et al., 2002; Björklund & Jäntti, 1997). Lee & Solon (2009) find that intergenerational income elasticities have been relatively stable over recent decades in the United States while Aaronson & Mazumder (2008) estimate increases in intergenerational income elasticities over the earlier period of 1940 to 1980.[2]

The focus of these studies has largely been on the relationship between sons and fathers. The relationship between the economic outcomes of daughters and their parents has received far less attention. As Chadwick & Solon (2002) note, "This neglect of daughters has stemmed partly from unconscious sexism and partly from a recognition that, in a society in which married women's labor-force participation rates are lower than men's, women's earnings may often be an unreliable indicator of their economic status." This second issue, the reliability of earnings as an indicator of economic status for females, is one of the major obstacles to addressing mobility differences

[1] For an alternative approach using wealth rather than income, see Steckel & Krishnan (1992).

[2] Studies of income mobility in the United States have been restricted to the period from 1940 onward as the 1940 federal census was the first to collect income data.

across genders. It is unclear how one should interpret observed earnings if work decisions are not based solely on maximizing an individual's income, an issue for both genders but far more pronounced for females.

The prevailing approach to extending mobility estimates to females has been to use measures of household income rather than daughter's income to estimate elasticities. Chadwick & Solon (2002) follow this approach for the United States and find that the intergenerational correlation in earnings is somewhat weaker for females than males but still strong overall. Lee & Solon (2009) estimate separate intergenerational income elasticities for sons and daughters between 1981 and 2000 and find elasticities for both groups to be similar in magnitude. While this approach to estimating female mobility rates is a major improvement over ignoring the existence of females, it does have its limitations. In particular, the degree to which household income measures the economic outcome of a particular individual, male or female, depends on the intrahousehold distribution of resources. If the distribution of household resources is changing from one generation to the next, either due to changes in the relative status of household members or changes in the relative contributions of household members to overall earnings, household earnings may have a very different interpretation in terms of economic success across generations and genders. These issues would become more pronounced if these studies were extended to earlier time periods in which the structure of the household and the allocation of labor by gender went through dramatic changes.

While intergenerational income elasticities are the standard for studies of modern mobility rates, they cannot typically be used for historical studies due to a lack of income data. The primary approach to measuring historical mobility rates has been focus on measures of occupational mobility by linking fathers and sons across federal censuses. The ability to match an individual across censuses circumvents the problem of not having historical longitudinal studies. However, these federal censuses contain a very limited set of economic variables, typically just employment status and occupation. Income first appears in the 1940 census federal which has only recently become public. Consequently, estimating historical intergenerational income elasticities which require multiple generations of income data has not been possible.[3]

[3] One exception to this is the 1915 Iowa state census which did include annual earnings. Parman (2011) uses a linked father-son sample based on this census to demonstrate that intergenerational income elasticities in Iowa in 1915 were much lower than modern intergenerational income elasticities. More recently, Feigenbaum (2015) has exploited the

Instead, studies have attempted to compare the relationship between a son's occupation and that of his father, interpreting a greater likelihood of transitions between occupational categories as evidence of greater mobility. An early study to take this approach was Guest et al. (1989). Ferrie (2005) and Long & Ferrie (2007) have undertaken substantially larger projects of matching sons and fathers over time and have found that occupational mobility rates were high in the mid-nineteenth century United States and have fallen over time, converging to the more modest modern levels. A major drawback to this approach is that occupational transitions are difficult to map into changes in welfare. Observing an occupational transition does not indicate whether the economic status of a son has improved or declined. Further complicating interpretation is the fact that the variation in income and wealth within occupations can be quite substantial suggesting that even in the absence of occupational mobility there may still be substantial income or wealth mobility.

These limitations of occupational mobility are compounded if one is interested in gender differences in mobility. With the labor force participation and occupational distribution of females changing dramatically over the past two centuries, the interpretation of female occupational mobility would change substantially over time. An alternative would be to consider how the occupation of a daughter's spouse compares to the occupation of her father. While this may give more meaningful measures over time, the estimation of this form of occupational mobility is thwarted by a more basic problem: the adoption of the husband's surname at the time of marriage. Matching individuals across censuses is done by identifying the son in the census as an adult and then searching for him in his childhood household. This is possible because his full name on the census will be the same when he is a child and when he is an adult. This will not be the case for females. Maiden name is not included on the census for married females. Consequently, a married female could be matched to her childhood household solely on the basis of first name and birth year. This is an insufficient amount of information for uniquely identifying the female as a child.[4]

release of the 1940 federal census data to estimate intergenerational income elasticities using these 1915 Iowa incomes for fathers and incomes reported in the 1940 census for sons.

[4]Olivetti & Paserman (2015) circumvent this restriction by creating pseudo-links across generations, using given first names to predict the socioeconomic status of a daughter's parents. A limitation of this pseudo-link approach is that it does not allow you to control for childhood household characteristics in any analysis.

3 Using Health Outcomes to Estimate Mobility

This paper takes an alternative approach to estimating historical mobility rates that addresses several of the issues with both measuring and interpreting mobility over time and across genders. The basic approach is to use long run health outcomes, specifically longevity and longevity adjusted for years in poor health. This approach offers distinct advantages over both income and occupational mobility. The first is in interpretation. An improvement in longevity can be directly interpreted as an improvement in welfare, whether for a female or male. Unlike household income or wealth, longevity is specific to the individual being measured allowing for us to draw meaningful distinctions between the outcomes of husbands and wives without any assumptions about the inner workings of the household. Unlike occupational transitions, improved longevity clearly implies improved welfare whether one is considering the nineteenth century or the twentieth century. Thus long run health outcomes offer a way to measure intergenerational mobility rates that can be compared meaningfully across time and gender in ways that occupational, income or wealth mobility rates cannot.

The second advantage of relying on health outcomes is that it solves several of the data limitations suffered by occupational and income mobility measures. By obtaining long run health outcomes from death certificates, the details of which will be discussed in the next section, several of the constraints for both modern and historical mobility studies can be avoided. Compared to modern studies relying on longitudinal surveys, death certificates offer a larger sample of individuals to study that covers a longer time period: death certificates have been maintained for complete state populations in many cases back to the late nineteenth century. Relative to historical studies, death certificates have the distinct advantage of giving both the married and maiden names of females, allowing adult females to be matched to their parents.

The estimation of intergenerational correlations in health outcomes has received attention in the past, largely in the demography and epidemiology literatures. Attempts to estimate correlations in longevity across generations have been severely limited by data availability. Modern health surveys have typically not been in place long enough to observe the death of both the participants and their parents. Several studies have attempted instead to use family histories of royalty or landed elites (see for example Beeton & Pearson (1899) and Gavrilova et al. (1998)).

However, one imagines that that the experiences of a particular line of royalty offer little insight into the mobility patterns of the population as a whole. Other approaches have relied on observing the outcomes of particularly long-lived individuals, typically either nonagenarians or centenarians (Abbot et al., 1978; Pearl & Pearl, 1934). However, once again the experience of a group that is by definition a set of outliers is not a particularly useful way to gauge the experience of the population.[5]

An alternative approach in the literature on intergenerational correlations in health has been to choose variables that are easier to observe. An example is the work of Classen (2009) on the intergenerational transmission of body mass index. Other studies tend to use different measures for the health of parents and the health of the children, allowing for estimates of intergenerational correlations in health to be obtained from cross-sectional data (see for example Bhalotra & Rawlings (2008) in which the authors estimate correlations between maternal adult height and a child's infant mortality risk). While these are reasonable approaches given modern health survey data, they are not feasible approaches to estimating historical intergenerational health correlations. The necessary health variables are simply unavailable.

By applying the census matching techniques of the occupational mobility literature to death certificates, it is possible to construct measures of intergenerational correlations in health that use comparable health measures for parents and children that can be observed for over a century for a representative sample of the population. This allows for extending modern estimates of correlations in health across generations back in time while avoiding the sample selection issues facing prior longevity studies. The following section describes the approach to creating a new intergenerational dataset of health data.

4 Data

4.1 Data Sources

To estimate historical correlations in health outcomes across generations, it is necessary to have a data source that provides a sufficient level of detailed health information for individuals, includes

[5]There is an additional strand of the demography literature concerned with the inheritance of longevity that compares the life spans of twins or adopted children to determine what component of longevity is genetic (see for example Yashin & Iachine (1997) and Sørensen et al. (1988)).

individual characteristics that can be used to match individuals to their parents, and covers a sufficiently large percentage of the population over a long enough period of time to observe both children and their parents and ideally to observe multiple generations of children and their parents. Death certificates provide a data source that satisfies all of these criteria.

The collection and preservation of detailed information on deceased individuals has been carried out at the state level in the United States. Many states started maintaining death certificates in their modern form as early as the late nineteenth century. Prior to individual death certificates, death information was often recorded as lists of deceased individuals. These earlier death rolls do not contain sufficient information to create decent measures of health outcomes or to match individuals to their parents. However, the death certificates that came into use in the late 1800s and early 1900s for a handful of states have consistently contained a remarkable level of detail. In most cases, these death certificates have been preserved on microfilm by state boards of health. In recent years, genealogy services have made efforts to create electronic indices of these death certificates searchable by basic characteristics such as name, date of birth and date of death.

This study utilizes death certificates from North Carolina. The choice of North Carolina is based on the availability and quality of death certificates. All death certificates for North Carolina from 1909 to 1975 have been scanned and indexed. This means that any individual dying in North Carolina over a 66 year period can be searched for and, if found, an image of her original death certificate can be downloaded. Beyond the availability of the North Carolina death certificates, they stand out for the quality of information they contain and the consistency of the information over time. Many other states contain only minimal information on their earliest death certificates. For North Carolina, even the earliest death certificates contain information on date of birth, date of death, place of birth, place of death, primary and contributing causes of death, other significant medical conditions, spouse's name and parents' names. The death certificates also contain socioeconomic status information in the form of occupation and educational attainment.[6] For a representative sample of the death certificates, see Figure 1.

The information on date of birth, date of death, causes of death and other medical conditions offers several ways of measuring health outcomes. The most direct measure is simply longevity

[6]For a complete listing of the variables reported by year of death certificate, see the appendix.

based on age at death. This longevity measure can also be adjusted by the reported duration of primary and contributory causes of death as well as other conditions to create a measure of healthy life span. Finally, the causes of death and other significant medical conditions are themselves interesting measures of health outcomes. For the initial analysis presented in this paper, I will focus on measures of longevity. As the sample is expanded and additional variables are transcribed, the scope of the analysis will be expanded to these additional measures of health outcomes.

The information on age and parents' names makes it possible to link individuals to their parents' household in the federal census when they were children. Information on the parents from the census can then be used to find the parents' death certificates allowing one to match these long term health outcomes of children to the exact same measures for their parents. The availability of the parents' names means that this linking process can be completed for females even if a female took her husband's surname. Federal censuses from 1860 through 1930 are used to establish the childhood households of individuals from the death certificates. While these censuses vary in format over the years, they all contain the name, birth state, age and occupation data that are relevant to this study.

4.2 Constructing an Intergenerational Dataset

The construction of the intergenerational dataset begins with a sample of death certificates for individuals who died between the years of 1934 and 1974. The upper end of this range is determined by the availability of digitized death certificates. While individual death certificates are publicly available for years after 1974, they are not electronically indexed in a way that allows for searching for a large number of individuals. The lower end of the range is chosen such that most individuals will have parents that died after 1909, the first year for which we have indexed, complete death certificates. The initial sample is drawn entirely from Mecklenburg County. Mecklenburg is chosen because it contains the city of Charlotte, providing a far more heterogeneous sample in terms of rural and urban places of residence, occupational distribution and race than if a more rural county was chosen. Work is currently under way to expand the data to include a random sample of all North Carolina counties.[7]

To create a sample of individuals for a particular year, I first use a computer script that extracts all of the transcribed data from the death certificates from that year. The transcribed data includes full name, gender, race, age at death, birth date, birth place (city, state, country), death date, death location (city, county), spouse's name, father's name and mother's name. With all of these data extracted to a single dataset, it is then possible to sample the individuals conditioning on any of these variables, oversampling any particular groups of interest. The preliminary dataset in this paper uses a ten percent sample of the Mecklenburg death certificates beginning with 1934.

It would be possible to search for the parents directly in the death certificates using the mother and father's name variables. However, without any additional information, it is rarely possible to identify unique matches in the death certificate records. What is required to have any success with matching parents to the death certificates is additional information on parents' ages and birthplaces. This is information that can be obtained by first matching individuals to their childhood households in the federal census.

To link to the federal census, I search by name, birth year and birth state in the earliest federal census for which an individual was alive (if an individual's birth year was during a census year, the next census is used). Choosing the earliest possible census affords the best chance to find the individual still living in her parents' household. The parents' names given on the death certificate are used to confirm census matches. Once an individual is matched to the federal census, several pieces of data are transcribed from an image of the original census record. These include household size, household location, and the parents' ages, birthplaces and occupations.

The additional variables from the census provide sufficient information to uniquely match the parents to their death certificates. I search the death certificates for each parent using the parent's name, year of birth and state of birth. Spouse's name is used to narrow the search results if multiple matches are returned. Once a parent is successfully matched to the death certificate records, the information from both the parent's death certificate and the child's death certificate are transcribed from images of the original certificates. This adds information on occupation,

[7]A random one percent sample of all death certificates for the years 1934 to 1974 for all counties has been linked. Cleaning and coding of variables in this sample is in progress.

exact birth date, death date, cause of death and other significant conditions to the dataset for both the child and parent.

4.3 Linkage Success Rates and Selection Bias

As with any study linking individuals across historical documents, failure to to link a subset of individuals is an issue. A substantial number of individuals drop out of the dataset as a result of not being found in the federal census or their parents not being found in the death certificate records. Despite the loss of a large percentage of individuals, a reasonable sample size can still be obtained because of the availability of all of the death records for North Carolina. However, the large fraction of individuals that are lost does lead to concerns of sample selection bias. Of particular concern are differences by gender and by race in the likelihood of individuals being successfully matched. For gender, we may expect differences in linking success because of the fundamentally different information available for males and females, in particular the changes in a female's last name. For race, there is a large literature concerned with the accuracy of historical data for black individuals (see for example Elo & Preston (1994); Elo et al. (1996)). If matching rates differ substantially across gender or across race, it raises concerns that any differences in observed intergenerational mobility between these groups may be driven by selection issues rather than actual differences in mobility.

Table 1 provides summary statistics for the success rates at each stage of the linking process by gender and race. At the first stage, failure to find individuals in the census is a result of both difficulty in finding individuals in their parents' household stemming from misspellings, poor enumerator handwriting and common names as well as a subset of the individuals being born after 1930. 18 percent of the individuals cannot be matched to the federal census because they were born after 1930, the most recent publicly available federal census at the time the pilot dataset was constructed.[8] As a result, the final linked sample is underrepresentative of individuals from the more recent birth cohorts dying at young ages. This problem is more severe for the black population than the white population (25 percent of black individuals in the original sample were born after 1930).

[8]Federal censuses are released to the public after 72 years. The 1940 federal census is now publicly available and is used for the linking of the statewide one percent sample. The availability of the 1940 census substantially reduces the number of individuals dropped from the data set for dying at a young age.

Of those individuals who were born early enough to be found in the federal census, 40 percent could be successfully matched to their childhood households in the census. While this is a rather good match rate for historical data, it should be noted that the high match rate is being driven by the white subset of the sample. Nearly 45 percent of white individuals born before 1930 were successfully matched to the federal census while only 20 percent of the black individuals born before 1930 were found in the census. While racial differences in the match rates are quite pronounced, there are no significant differences between the match rates for males and females.

The racial differences in the success rates are compounded by the second stage of linking in which the parents are linked to their death certificate records. Of those individuals who were successfully matched to the federal census, 50 percent had at least one parent successfully matched to the death certificate records and under 20 percent had both parents successfully matched to the death certificates. Once again, the success rates are substantially higher for whites than blacks but similar across genders, both in terms of the gender of the child and the gender of the parent.

If the failure to match is random, these low match rates from the intial sample to the final set of observations with death certificate information for the parents would not be a concern. The white individuals in the sample could be treated as a random sample of the white population and an oversample of the black population could be used to obtain sufficiently large sample sizes to estimate racial differences in mobility. The similarities in match rates across genders would suggest that there are no obvious differences in sample selection issues for males and females that would necessitate different sampling strategies by gender. Certainly some reasons for match failures are random (poor census enumerator handwriting, common first names, etc.). However, a variety of other reasons for match failures are not random and could lead to major sample selection bias issues. More geographically mobile individuals will be harder to match. People more prone to misstate their age will be harder to find. Individuals from single parent households will be more difficult to find as there are fewer identifying pieces of information to work with. Table 2 presents summary statistics by gender, race and matching success to assess the extent of these selection issues.

Differences in the mean characteristics of matched and unmatched individuals do raise concerns about sample selection. While age at death is not strongly correlated with the probability of being matched for white individuals, the average age at death is far lower for black individuals who could not be matched to the census than those who could be matched. For all of the other observable characteristics, whites and blacks exhibit similar patterns. Individuals who could be matched appear to be less geographically mobile than people who could not be matched, with the proportion of people born in North Carolina being larger among matched individuals than unmatched individuals. One of the largest differences between matched individuals and unmatched individuals is the proportion listing no occupation, with a far larger percentage of matched individuals listing an occupation. This is unsurprising given that individuals with missing information for occupation often had other information missing from the death certificate, making them more difficult to match to census records.

These various linking statistics leave us with several issues in terms of the representativeness of the sample. Linking success does not seem to be gender dependent giving us no reason to worry that the daughters in the sample are any more or less representative of the female population than the sons in the sample are of the male population. However, the low match rates for the black population and the summary statistics by match success do suggest that there are several important ways in which the sample will be unrepresentative of the population as a whole. In its current state, the sample is underpresentative of blacks and particularly blacks dying at young ages, individuals born in more recent cohorts who died early, and geographically mobile individuals. To the extent that intergenerational correlations in health differ by any of these characteristics, the generalizability of any results to the population as a whole will be limited.

4.4 Age Reporting Issues

With the availability of birth year from the death certificates and age from the federal census when an individual was a child, it is possible to assess how accurately age is reported. A major concern when using these data to study longevity is that self-reported age is a noisy measure of true age (see Mason & Cope (1987) for an extensive discussion of various sources of age misreporting and Rosenwaike & Logue (1983) and Hill et al. (2000) on the accuracy of reported ages particularly among older individuals). With noisy measures of longevity, estimates of the intergenerational correlation in longevity will suffer from an attenuation bias, the severity of which will depend on the frequency and magnitude of age reporting errors. Age misreporting creates an additional source of bias by making incorrect matches between the census and death certificates more likely. Including incorrectly matched individuals in the dataset will lead to underestimates of the strength of the relationship between parents' and children's outcomes, as these outcomes will be uncorrelated for incorrectly paired parents and children.

A basic way to check for age reporting issues is to compare the birth year given on a person's death certificate to the birth year implied by the age given on the census when the person was a child. If ages are being reported accurately, these two birth years should be in agreement (possibly differing by one year depending on whether the census was administered before or after the person's birthday). Table 3 summarizes the mean difference between the birth year stated on the death certificate and birth year implied by age given on the census by gender and race. Age misreporting is certainly present for whites but relatively mild and similar across genders. The negative mean for the difference in birth years is primarily a result of the instances in which the census is administered prior to a person's birthday. The average magnitude of the difference in birth years is less than one year for the children in the sample and less than two years for parents.[9] The picture is quite different for blacks in the sample. It appears that age is systematically underreported on the death certificate and by a fairly large amount. This finding is consistent with the existing literature on age misreporting and race (Elo & Preston (1994); Elo et al. (1996)).

These results suggest that there will be potentially large measurement error issues when calculating longevity, a major concern when trying to estimate the relationship between a child's life span and her parent's life span. A possible way to reduce the measurement error is to determine whether the birth year on the death certificate or the birth year implied by the census is more reliable. If one is substantially more reliable than the other, then it would be sensible to use the more reliable one in the analysis. If they are equally reliable, or rather equally unreliable, one can attempt to eliminate the measurement error bias by instrumenting for one noisily measured birth year with the other in the manner of Ashenfelter & Krueger (1994).[10]

[9]Note that in other studies linking individuals across censuses, large age discrepancies will not be present because individuals with large age discrepancies are discarded as poor matches. However, this is a feature of how individuals are matched across censuses and not indicative of a lack of age misreporting. The reason age misreporting is observable in this study is that individuals can also be matched on parents' names, making it possible to uniquely identify a match even when age is misreported.

[10]This approach depends on the measurement error being orthogonal to the true birth year. If it is not, the IV estimates will still be biased (Black et al., 2000).

One approach to determining which birth year is more reliable is to construct measures of age heaping for both. Assuming that the probability of being born in a year ending in zero or five is the same as the probability of being born in a year ending in any other digit, roughly 20 percent of the individuals should have death certificate birth years ending in either a zero or five (or ages ending in a zero or five in the case of the census). If individuals tend to round their ages due to uncertainty about their true age, we would expect them to round to zeros and fives, or at least round to zeros and fives more frequently than other digits. A greater percentage of individuals with birth years or ages ending in zeros or fives would indicate age is more likely to be reported with noise.

Table 4 examines this measure of age heaping for the birth year stated on the death certificate and the ages given on the federal census by gender and race for the linked sample of children and parents.[11] There is no evidence of age heaping for children, either on the census or the death certificate. It does appear that there is age heaping for both fathers and mothers in the census. Once again, these age reporting issues are more pronounced for the black observations relative to the white observations. The parents' death certificate birth years do not show evidence of age heaping. For this reason, using death certificate birth years will be the preferred approach when calculating life spans. Results are also presented using the life span based on the census birth year to instrument for life span based on the death certificate birth year. However, there is a strong likelihood that the age misreporting errors are negatively correlated with the true age at death which leads to the IV estimates being biased.

The extent of age misreporting and age heaping present in Table 3 and Table 4 raise concerns of selection bias in addition to the concerns of measurement error discussed above. If the likelihood of misreporting age is correlated with individuals characteristics related to longevity, the failure to match individuals with grossly misreported ages will potentially bias results by leading to an unrepresentative sample. To get a better sense of who tends to misreport ages, I use the Integrated Public Use Microdata Series (IPUMS) samples of the 1880, 1900, 1910, 1920 and 1930 federal censuses and the full set of North Carolina death certificates to get a fuller picture of

[11] It is not entirely clear whether the death certificate birth year or the age given on the death certificate should be used to measure age heaping. I assume that the date of birth is reported by the person present at time of death. In this case, the birth year would be the appropriate variable to check for age heaping. However, it is possible that the person present reports the age and then birth year is imputed from the age in which case the age heaping would be observed in the distribution of the final digit of age rather than birth year.

which subsets of the population tend to round their ages and therefore which are most likely to be underrepresented in the linked dataset. Figure 2 shows the distribution of the last digit of the reported ages for all adult North Carolina residents in the 1% IPUMS federal census samples over the age of 19. While there is slight evidence of age heaping for white males and females, there is pronounced age heaping for both black males and black females with significantly higher proportions of zeros and ones than other digits. Turning to reported ages for children between the ages of 10 and 19, Figure 3 reveals a different age heaping problem. There is no noticeable heaping on zeros and fives. However, even ages are reported with substantially higher frequency than odd ages, particularly for black individuals revealing that even at young ages, age misreporting is problematic. Turning to all adults in the North Carolina death certificates, Figure 4, it is apparent that age heaping is not restricted to the census. While white individuals show no signs of age heaping on the death certificates, zeros and fives are once again overrepresented in the reported birth years of black individuals.

The IPUMS census samples provide a wealth of other personal characteristics to assess what the correlates of age heaping are at an individual level. Table 5 and Table 6 present regressions of proxies of age heaping on various individual characteristics. For children and adults, one proxy for age heaping is simply an indicator variable for whether the individual's age ends in a zero or five. Given the patterns in Figure 3, a second proxy is used for children, an indicator variable equal to one if the child's age ends in an even number. For all of the regressions, a positive coefficient can be interpreted as an increase in that variable being associated with greater levels of age heaping and therefore a higher likelihood of age misreporting. In the regressions for children, few of the independent variables end up having significant correlations with age heaping. However, in the adult regressions race is strongly related to age heaping, with black individuals more likely to report an age ending in zero or five than white individuals. This effect is more pronounced for black females relative to white females than black males relative to white males. Individuals living on farms are less likely to exhibit age heaping. Unsurprisingly, higher incomes and higher levels of literacy are associated with lower levels of age heaping. Taken as a whole, the regressions suggest that age misreporting will create sample selection biases arising from black and low-socioeconomic status individuals being underrepresented in the linked sample.

5 Estimates of Intergenerational Correlations in Health

5.1 The Relationship Between Occupation, Longevity and Mobility

Before turning to measures of mobility based on intergenerational health correlations, it is instructive to compare the health measures to more traditional measures of occupational mobility. Comparing longevity to occupations establishes that measures of longevity are related to socioeconomic status and offers guidance as to how health-based estimates of mobility relate to the existing literature on intergenerational occupational mobility.

The occupation data come from the occupation question on the death certificates. I focus on the occupations of sons and fathers only as the vast majority of females list either housework or housewife as occupation. Occupation on the death certificate is supposed to be the individual's usual occupation defined as, "the type of work done during most of working life."[12] I group occupations according to the classification system used by Ferrie (2005), creating four categories of occupations: farmer, white collar, skilled or semi-skilled and unskilled.

Table 7 shows the distribution of occupations for all males in the sample by birth cohort. The changing occupational distribution suggests that there was substantial occupational mobility taking place over the 1800s and early 1900s as a result of structural change in the economy. The percentage of farmers dropped significantly over this period while the percentage of skilled and white collar workers both rose significantly. The occupational mobility driven by both this structural change in the economy as well as more general mobility is shown in Table 8 showing the occupational transitions of sons relative to their fathers. What emerges from Table 8 is a picture of substantial but not complete occupational mobility. A significant number of occupational transitions are observed but there is occupational persistence from one generation to the next, particularly for farmers and white collar workers.

Given the death certificate information, it is possible to ask whether long term health outcomes differed by occupational category. In particular, one can assess whether the mean life span

[12] This is different than the census occupations used in other mobility studies as the census gives the individual's current occupation. If occupations change over the lifetime, an intergenerational mobility estimate based on census occupation will be higher than a mobility estimate based on usual occupation over one's lifetime, the occupation measure available from the death certificates.

of an individual differed across occupations both in terms of the son's occupational category and the father's occupational category. Table 9 gives mean life span for sons by the son's occupational category and mean life span for both sons and daughters by father's occupational category. The differences suggest that there are strong links not only between one's own occupation and health status but also between father's occupation and the health status of his children, both male and female. The most striking feature of the data is the remarkably short life span for unskilled workers and their children. Sons with an unskilled occupation have a mean life span that is 8.5 years shorter than the mean life spans for any other occupational category. While the gap narrows when considering son's life span by father's occupation it is still the unskilled occupations that are associated with the shortest life spans. The mean life spans for daughters by father's occupation demonstrate the same patterns as those for the sons: the children of farmers have the longest life spans while the children of unskilled workers have the shortest life spans.

The interpretation of these differences is clouded by the difficulty in ranking these broad occupational categories by socioeconomic status. A familiar problem in the occupational mobility literature is that these broad occupational categories do not lend themselves to strict ordering in terms of income, wealth or other socioeconomic measures. For example, farmers as a group often contained some of the richest and poorest individuals in a community.[13] The large variances in life spans within occupational categories suggest that this overlap in outcomes across categories extends to health as well as income and wealth. It is therefore difficult to assess whether an occupational transition represents a change in welfare. The one reasonable exception to this is the unskilled category. The unskilled occupations are clearly at the bottom of the occupational ladder in terms of status and pay. While it is difficult to rank farmers, skilled or semi-skilled workers and white collar workers, it is more reasonable to assume that all three typically rank above unskilled workers. This is confirmed by the average age at death by occupational category.

Table 10 presents a slightly more formal approach to assessing the impact of occupational category on longevity. The table presents regression coefficients from OLS regressions of a child's life span on dummy variables for the father's occupational category and, in the case of sons, the

[13] A further complication with farmers is that they are the group most likely to live in a rural rather than an urban area. For the time period of this study, there is still a large health penalty to living in an urban area. The long life spans observed for farmers are almost certainly due in part to better living conditions in rural areas relative to urban areas in addition to whatever the average differences in economic outcomes are for farmers.

son's occupational categories. A quadratic in child's birth year is included to control for general changes in longevity over time that would otherwise get picked up by the occupational dummies (since occupational structure is changing over time as well). The results confirm the significant relationship between an individual's longevity and both his own occupation and that of his father.

These basic summary statistics of the relationship between occupational status and life span suggest that life span is capturing important aspects of socioeconomic status. In particular, unskilled occupations are associated with shorter life spans. This relationship exists both between a son's occupation and his life span as well as across generations, with both the sons and daughters of unskilled workers having shorter life spans on average than the sons and daughters of farmers, skilled or semi-skilled workers and white collar workers.

5.2 Health-based Intergenerational Mobility Estimates

I now turn to estimating correlations in health across generations. Figure 5 provides a picture of changes in longevity over time for males and females. In both figures, the sample is divided into individuals with parents who had longer life spans than the predicted life span based on their birth year and individuals with parents who had shorter life spans that the predicted life span based on their birth year. For both sons and daughters, it is clear that throughout the past century individuals with long-lived parents tended to have longer life spans than individuals with short-lived parents. However, the figures suggest that there may be differences across genders in the strength of this relationship between parent and child longevity. The gap in the curves is substantially larger throughout history for males compared to females and far more persistent through the second half of the century. The remainder of this section will attempt to quantify this relationship between the longevity of children and their parents.

To estimate the relationship in longevity across generations, I follow the approach taken by the literature on modern income mobility rates and calculate intergenerational elasticities in the variable of interest. In its most basic form, this entails regressing the log of a child's life span on the log of the parent's life span. However, there are several potential pitfalls in this approach when applied to longevity. The first is that we need to take into account the secular trends in lifespan. As Figure 5 confirms, average life span has been increasing over time. If we do not control for the time period in which a child is born, we would estimate a positive relationship between child's life span and parent's life span regardless of whether there is any transmission of health status from parent to child. The reason is simply that both the child's life span and parent's life span for a recent child-parent pair will tend to be larger than the life spans for a child-parent pair from an earlier cohort due to the secular trends in life span. Even if there is no relationship between a child's life span and parent's life span in a cross-section of a single cohort, we will find a positive relationship in a sample spanning multiple cohorts. It is therefore necessary to control for cohort in the regressions by including a quadratic in child's birth year. The model being estimated is then

$$\ln(\text{lifespan}_i^s) = \beta_0 + \beta_1 \ln(\text{lifespan}_i^f) + \beta_2 t_i + \beta_3 t_i^2 + \varepsilon_i \tag{1}$$

where lifespan_i^s is the life span of son i, lifespan_i^f is the life span of his father, t_i is the birth year of son i and ε_i is an error term of mean zero. The intergenerational longevity elasticity is given by the coefficient β_1. In an alternative specification, I also include an interaction term between child's birth year and the log of parental life span to capture changes in the intergenerational correlations in longevity over time.

Regression estimates are presented in Table 9. Separate regressions are run for females and males. To use the largest sample sizes possible, only the life span of the parent of the same gender as the child is used. This both retains those observations for which only one parent is observed and makes interpretation of the longevity elasticity simpler by comparing male life spans to male life spans in the son regressions and female life spans to female life spans in the daughter regressions. Regressions including both parents' life spans are included in Table 12.

Under every specification in Table 9, the coefficient on parent's log life span is large and highly significant. The magnitudes suggest that the intergenerational elasticity for life span is actually quite similar in magnitude to intergenerational income elasticities. Just as studies of income mobility have found that a ten percent increase in parents' income is associated with a roughly three percent increase in the child's income, the longevity estimates suggest that a ten percent increase in a parent's life span is associated with a two to three percent increase in a child's life span. The point estimates of the longevity elasticities are larger for sons than for daughters but these differences are insignificant given the standard errors of the coefficients.

It is worth remembering that the measurement error in ages may be biasing these coefficients toward zero. The actual relationship between child and parent longevity is likely stronger than the estimates suggest.

One concern when interpreting the results is censoring of both child and parent life spans. While the data do begin as a random sample of the universe of deaths, individuals in the later death cohorts whose parents outlive them will drop out of the sample if the parents have not died by 1974 and individuals whose parents died prior to 1909 will also drop out of the sample. This latter issue is particularly problematic for members of the earlier death cohorts who either had very long lives themselves or had parents with particularly short lives. To assess the biases introduced by truncating the available death years, I simulated father-son pairs, drawing the fathers' birth years from a uniform distribution and life spans from a distribution matched to the true population, generating sons for these fathers whose birth years relative to their fathers were drawn the population distribution of child ages relative to parent ages, and then generating the sons' life spans based on the estimated values of the coefficients in equation (1).[14] I then estimated the longevity elasticity, varying the range of years for which death certificates are observed for the regression sample. The resulting elasticity estimates by year range are shown in Figure 6 (the true elasticity used to generate the data was 0.28). As the graph shows, the censoring resulting from truncating the sample of death certificates leads to a downward bias on the elasticity estimate that becomes worse as the year range narrows. These simulations suggest that the elasticities in Table 9 are underestimates of the true intergenerational longevity elasticity.

The interaction term between child's birth year and the log of parent's life span does point to differences in the mobility patterns of males and females over time. For sons, the interaction term coefficient is positive and statistically significant at a 10 percent significance level, suggesting that the correlation in longevity between sons and fathers has strengthened over time. The magnitude is quite large: given the size of the coefficient, a son born in 1890 would have an intergenerational longevity elasticity of 0.12 compared to an elasticity of 0.31 for a son born in 1910. For daughters, the coefficient on the interaction term is actually negative but quite small

[14]The distributions of father life spans was based on all males in the universe of North Carolina death certificates. The distribution of sons' ages relative to father's ages was based on all sons of the head of household in the 1850 through 1960 IPUMS samples of the federal census.

and statistically insignificant revealing no evidence of a change in the correlation of longevity between mothers and daughters over the past century.

The regression results in Table 12 including both parents' life spans offer additional insight into the nature of these intergenerational correlations in longevity and highlight an additional difference between males and females. For sons, both the log of father's life span and the log of mother's life span have statistically significant and large coefficients. The coefficient for the log of father's life span is substantially larger than the coefficient on the log of mother's life span. For daughters this pattern in reversed, with the coefficient on the log of mother's life span being substantially larger than the coefficient on the log of father's life span. These results suggest that the transmission of long term health outcomes, at least in the case of longevity, is stronger between parents and children of the same gender but not necessarily limited to parents and children of the same gender. These findings are consistent with the work of Olivetti & Paserman (2015). Despite using a very different approach of creating synthetic links across cohorts and focusing on occupational mobility, Olivetti & Paserman find similar gender differences in mobility.

5.3 Removing the Genetic Component of Longevity

One concern unique to this focus on the intergenerational transmission of health as opposed to income or other measures of socioeconomic status is that the links in health may be primarily driven by genetics. While this does not render longevity uninteresting, it would make it difficult to use intergenerational longevity elasticities to measure changes in the strength of intergenerational correlations in socioeconomic status across gender and over time. Fortunately, the richness of the linked data offer an opportunity to test the extent to which the estimates are driven purely by genetics rather than socioeconomic status. In particular, the availability of information on both parents allows for estimating a longevity elasticity across spouses rather than across generations. Given that spouses do not share any genetics but do share a common environment, one that is strongly influenced by socioeconomic status, a strong correlation between the age at death of spouses would help support the notion that correlations across generations in longevity can be partially attributed to correlations in socioeconomic status across generations.

Table 13 presents regression results using the same specifications as the intergenerational longevity elasticity estimates but using one spouse's age at death as the dependent variable and the other spouse's age at death as the key independent variable. A quadratic in the birth year for the spouse serving as the dependent variable is included to once again control for general changes in longevity over time. The results reveal a remarkably strong relationship between spouses' life spans. The elasticity estimates of between 0.16 and 0.19 are similar in magnitude to the elasticities estimated across generations. These findings suggest that a large portion of the intergenerational correlations in longevity may be the product of socioeconomic status, not genetics.

5.4 Modern Comparisons

One of the key advantages of a historical study of longevity is the availability of data. Few comparable modern data sources exist and most health surveys focus on either early-life outcomes or health behaviors, not on mortality. Having data on the mortality of both an individual and his or her parents is even rarer. However, there are ongoing efforts to link past participants in modern health surveys to mortality data. In particular, the National Center for Health Statistics (NCHS) has been linking survey participants to the National Death Index (NDI), a centralized database of all U.S. deaths beginning in 1979. In 1987, the National Interview Health Survey asked respondents the age at death for each of their parents. For those individuals in the survey who were linked to the NDI and whose parents had died prior to 1987, it is possible to estimate an intergenerational longevity elasticity to provide a modern point of comparison for the North Carolina estimates.

To estimate the intergenerational longevity elasticities, the National Interview Health Survey data for 1987, the one year with the parents' age at death questions, is taken from the Integrated Health Interview Series (IHIS) public files. Given that mortality patterns will differ across locations, to make the best comparison for the North Carolina sample region fixed effects are included in all regressions (region is the most detailed information on location available through IHIS). Results of the intergenerational longevity elasticity regressions are presented in Table 14.[15] The estimates using modern data are dramatically different from those in Table 11 or Table 12, with the modern elasticities being far smaller in magnitude for both males and females. This

[15] Table 14 only reports results for white males and females. The sample sizes for black males and females were too small to yield precise estimates. Only 316 black males and 234 black females had ages at death reported for both parents.

stands in stark contrast to other studies of mobility over time that have found increases in the strength of the relationship between parents' and children's incomes and occupations and to the North Carolina estimates which reveal increasing elasticities over time.

One possible reason for these markedly smaller intergenerational longevity elasticities is an issue of sample selection. While the linked sample can take advantage of seven decades of death certificates, individuals only make it into the IHIS sample if they die between 1987 and 2011 (the most recent year for which the linking to the NDI has been completed) and if their parents died prior to 1987. This means that short-lived individuals will be more likely to make it into the sample. If these individuals are dying from more random causes such as accidents, we might expect to estimate a much weaker relationship between parents' and children's outcomes. To see whether this is indeed the case, Table 15 presents the intergenerational income elasticities estimated under varying cutoffs for the age at death for both the child and the parent. Indeed, when the samples are restricted to people dying at older ages and therefore more likely to be dying from natural causes, the elasticities increase in magnitude and in statistical significance. However, they remain an order of magnitude smaller than the North Carolina estimates.

To get a sense of what to make of these differences between the modern estimates and the historical North Carolina estimates, it is useful to return to the spouse regressions presented in Section 5.3. Repeating this exercise for the IHIS data produces results much more in line with the historical estimates, as seen in Table 16. The correlations between spouses' longevity remain strong and quite similar in magnitude to those found in the North Carolina data. This suggests that it is not necessarily a sample selection issue driving the low modern intergenerational longevity elasticities. Spouse's sharing no genetics but similar socioeconomic conditions have similar ages at death in modern data. What the results may suggest is that socioeconomic status later in life remains a strong predictor of mortality but that the socioeconomic gradient for early-life health inputs has declined over time. While this would be consistent with many of the advances in early-childhood health of the early twentieth century being available to all individuals, for example improvements in water quality, immunizations, and so on, better data and far more extensive analysis is needed before any conclusions can be reached.

6 Extensions

These initial results on the intergenerational correlations in life span are encouraging evidence that health measures may offer a practical approach to historical mobility rates that can make meaningful distinctions between the experiences of males and females. Currently the dataset is being extended both in terms of the geographical coverage of the sample and the density of the sample. A one percent sample of the entire universe of North Carolina death certificates has been linked and is now being transcribed. This substantially larger sample will allow for better estimates of the changes in intergenerational correlations in health over time. There are a variety of reasons to believe that changes in the strength of this relationship have been nonlinear over time. Better data coverage over time and space will allow for investigating how these intergenerational relationships changed with the expansion of the female labor sector, waves of rural to urban migration, changes in voting rights of females, and demographic transitions. Additionally, it will be possible to explore racial differences in the intergenerational transmission of health status with a larger sample.

A second extension currently under way is the utilization of the rich medical information given on the death certificates. The majority of death certificates give detailed information about the cause of death as well as other significant medical conditions. Included in this information are intervals between onset and death for the various conditions. Many of the other significant medical conditions listed are chronic illnesses and ailments. This interval information makes it possible to adjust actual life spans by the number of years spent in poor health. Figure 7 depicts these adjusted healthy life spans relative to actual life spans for the current sample. This measure of healthy life spans would offer a measure of long run health that may be a better proxy for overall welfare than the actual life spans that are focus of this paper. Devising a way to categorize the various conditions and their impact on quality of life could further refine the measures of long term health and the estimates of correlations in those measures across generations.

One final promising direction to expand this research involves the incorporation of additional census information. By linking the sons and daughters of the sample to their adult households in the federal census, additional variables can be added to the dataset including spouse's occupation, age at marriage, number of children and their ages. These variables would allow for a fuller exploration of female mobility. Occupational mobility measures comparable to those used in Ferrie (2005) and Long & Ferrie (2007) could be constructed for females by comparing spouse's occupation to father's occupation. Additionally, the data would allow for identifying the determinants of female mobility including age at marriage, fertility patterns including age at first child and the number and spacing of children, labor force participation decisions and spouse characteristics.

7 Conclusion

This paper has introduced a new intergenerational dataset of death certificate data that allows for estimating intergenerational correlations in long run health outcomes. The unique features of name reporting on death certificates allows for extending the census linking strategies of other studies to females. The measures of health across generations constructed from the death certificates offer a new approach to measuring historical mobility rates that avoids several of the limitations of current measures. They provide a metric for welfare in the form of longevity that has a clear interpretation that is consistent across genders and over time.

The results suggest that longevity does provide a meaningful measure of welfare and that there have been persistent intergenerational correlations in longevity for both males and females over the past century. I find a positive correlation between a son's life span and his occupational status, with sons in unskilled jobs having a life span several years shorter on average than son's in skilled, semi-skilled, white collar or agricultural positions. This correlation between a son's life span and his occupational status also extends across generations, with both sons' and daughters' life spans being strongly correlated with their father's occupational status.

The estimates of intergenerational life span elasticities reveal strong correlations between parents' longevity and that of their children remarkably similar in magnitude to intergenerational income elasticity estimates for the United States. These correlations are stronger between parents and children of the same gender. The elasticities for daughters appear relatively stable over time while there is evidence of a gradual increase in the strength of the correlation between father and son life spans over the past century.

These results suggest that long run health outcomes are a promising new direction for studying historical mobility patterns. The data linking approach pursued in this paper can be extended to a more thorough analysis of mobility patterns across gender and race for the late nineteenth and early twentieth centuries. Combining the death certificate data introduced here with the additional data available through the federal census would allow for explaining the differences in male and female mobility over time through individual marriage choices, labor market participation and fertility patterns, offering a substantially more complete picture of the history of American mobility.

References

Aaberge, R., Bjorklund, A., Jantti, M., Palme, M., Pedersen, P., Smith, N., & Wennemo, T. (2002). Income inequality and income mobility in the Scandinavian countries compared to the United States. *Review of Income and Wealth*, *48*(4), 443–469.

Aaronson, D., & Mazumder, B. (2008). Intergenerational economic mobility in the united states, 1940 to 2000. *Journal of Human Resources*, *43*(1), 139–172.

Abbot, M., Abbey, H., Boiling, D., & Murphy, E. (1978). The familial component in longevity-a study of offspring of nonagenarians: 111 lntrafamilial studies. *American Journal of Medical Genetics*, *2*, 105–120.

Ashenfelter, O., & Krueger, A. (1994). Estimates of the economic return to schooling from a new sample of twins. *The American Economic Review*, *84*(5), 1157–1173.

Beeton, M., & Pearson, K. (1899). Data for the Problem of Evolution in Man. II. A First Study of the Inheritance of Longevity and the Selective Death-rate in Man. *Proceedings of the Royal society of London*, *65*, 290–305.

Bhalotra, S., & Rawlings, S. (2008). The Intergenerational Correlation of Health in Developing Countries. Tech. rep., mimeo, University of Bristol, UK.

Björklund, A., & Jäntti, M. (1997). Intergenerational income mobility in Sweden compared to the United States. *The American Economic Review*, (pp. 1009–1018).

Black, D. A., Berger, M. C., & Scott, F. A. (2000). Bounding parameter estimates with nonclassical measurement error. *Journal of the American Statistical Association*, *95*(451), 739–748.

Chadwick, L., & Solon, G. (2002). Intergenerational income mobility among daughters. *American Economic Review*, *92*(1), 335–344.

Classen, T. (2009). Measures of the intergenerational transmission of body mass index between mothers and their children in the United States, 1981-2004. *Economics & Human Biology*.

Elo, I., & Preston, S. (1994). Estimating African-American mortality from inaccurate data.

Demography, *31*(3), 427–458.

Elo, I., Preston, S., Rosenwaike, I., Hill, M., & Cheney, T. (1996). Consistency of age reporting on death certificates and social security administration records among elderly African Americans. *Social Science Research*, *25*, 292–307.

Ferrie, J. (2005). History lessons: The end of American exceptionalism? Mobility in the United States since 1850. *Journal of Economic Perspectives*, *19*(3), 199–215.

Gavrilova, N., Gavrilov, L., Evdokushkina, G., Semyonova, V., Gavrilova, A., Evdokushkina, N., Kushnareva, Y., Kroutko, V., & Andreyev, A. (1998). Evolution, mutations, and human longevity: European royal and noble families. *Human Biology*, *70*, 799–804.

Guest, A., Landale, N., & McCann, J. (1989). Intergenerational Occupational Mobility in the Late 19th Century United States. *Social Forces*, *68*(2), 351–378.

Hill, M., Preston, S., & Rosenwaike, I. (2000). Age reporting among white Americans aged 85+: results of a record linkage study. *Demography*, *37*(2), 175–186.

Lee, C., & Solon, G. (2009). Trends in intergenerational income mobility. *The Review of Economics and Statistics*, *91*(4), 766–772.

Long, J., & Ferrie, J. (2007). The path to convergence: intergenerational occupational mobility in Britain and the US in three eras. *Economic Journal*, *117*(519), 61.

Mason, K., & Cope, L. (1987). Sources of age and date-of-birth misreporting in the 1900 US census. *Demography*, *24*(4), 563–573.

North Carolina State Board of Health, Bureau of Vital Statistics (1909). *North Carolina Death Certificates, 1909-1975*. Raleigh, NC: North Carolina State Archives. Microfilm S.123 Rolls 19-242, 280, 313-682, 1040-1297.

North Carolina State Board of Health, Bureau of Vital Statistics (2007). *North Carolina Death Certificates, 1909-1975*. Provo, UT: Ancestry.com Operations Inc. On-line database accessed through Ancestry.com.

Olivetti, C., & Paserman, D. (2015). In the name of the son (and the daughter): Intergenerational mobility in the united states, 1850-1930. *American Economic Review*, *105*(8).

Parman, J. (2011). American mobility and the expansion of public education. *The Journal of Economic History*, *71*(01), 105–132.

Pearl, R., & Pearl, R. (1934). *The ancestry of the long-lived*. Johns Hopkins press.

Rosenwaike, I., & Logue, B. (1983). Accuracy of death certificate ages for the extreme aged. *Demography*, (pp. 569–585).

Solon, G. (1999). Intergenerational mobility in the labor market. *Handbook of labor economics*, *3*, 1761–1800.

Solon, G. (2002). Cross-country differences in intergenerational earnings mobility. *Journal of*

Economic Perspectives, *16*(3), 59–66.

Sørensen, T. I., Nielsen, G. G., Andersen, P. K., & Teasdale, T. W. (1988). Genetic and environmental influences on premature death in adult adoptees. *New England Journal of Medicine*, *318*(12), 727–732.

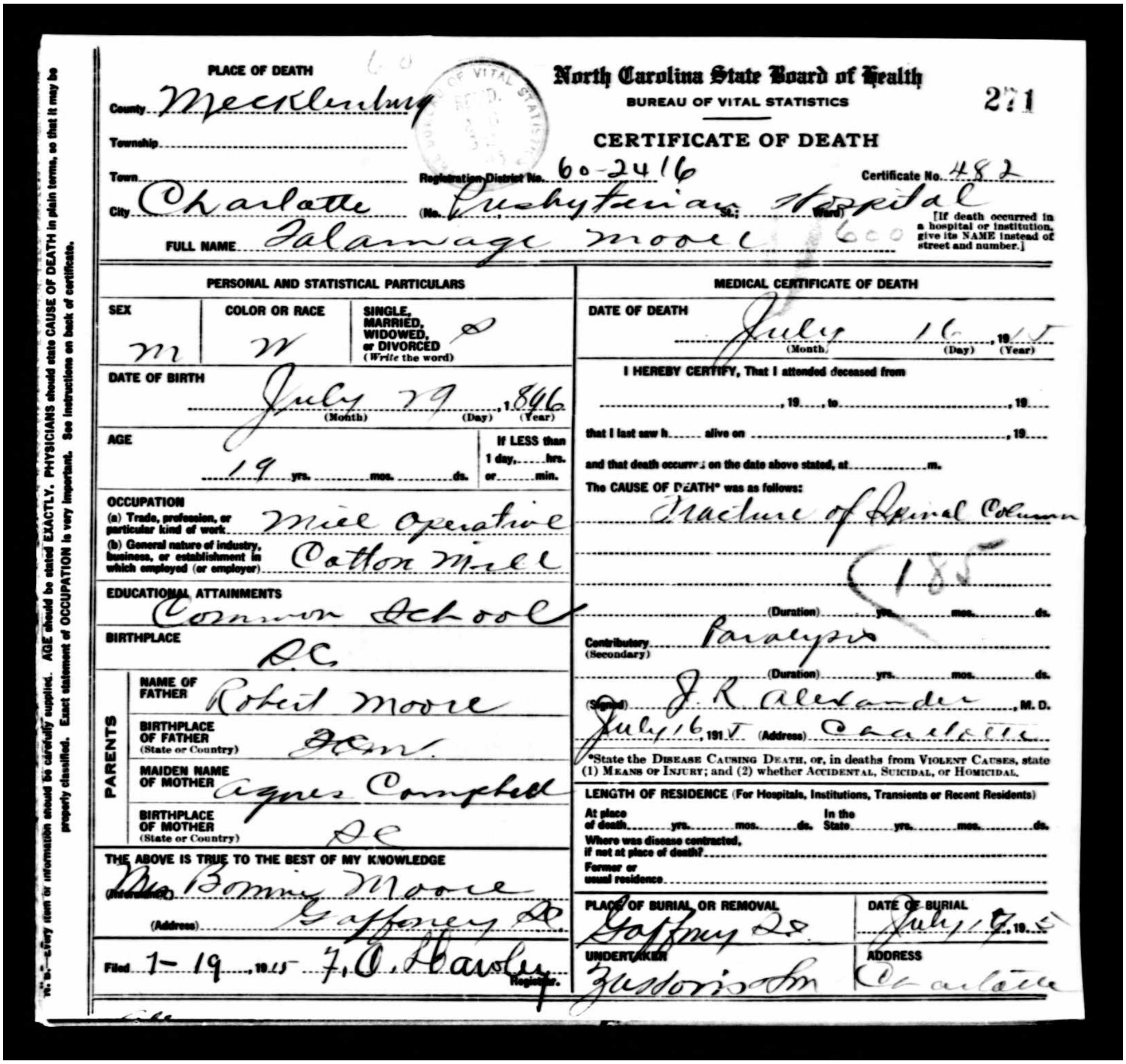
Steckel, R., & Krishnan, J. (1992). Wealth mobility in America: a view from the national longitudinal survey. *NBER Working Paper*.

United States of America, Bureau of the Census (1860). *Eigth through Fifteenth Census of the United States, 1860-1930*. Washington, D.C.: National Archives and Records Administration. Microfilm records at the National Archives and Records Administration.

United States of America, Bureau of the Census (2004). *Census of the United States, 1860-1930*. Provo, UT: Ancestry.com Operations Inc. On-line database accessed through Ancestry.com.

Yashin, A., & Iachine, I. (1997). How frailty models can be used for evaluating longevity limits: Taking advantage of an interdisciplinary approach. *Demography*, *34*(1), 31–48.

8 Tables and Figures



PLACE OF DEATH
County Mecklenburg
Township
Town
City Charlotte (No. Presbyterian Hospital St., Ward) [If death occurred in a hospital or institution, give its NAME instead of street and number.]
Registration District No. 60-2416
North Carolina State Board of Health
BUREAU OF VITAL STATISTICS
CERTIFICATE OF DEATH
271
Certificate No. 482
FULL NAME Talamage Moore

PERSONAL AND STATISTICAL PARTICULARS
SEX m
COLOR OR RACE w
SINGLE, MARRIED, WIDOWED, or DIVORCED (Write the word) S
DATE OF BIRTH July 29, 1896 (Month) (Day) (Year)
AGE 19 yrs. mos. ds. If LESS than 1 day, hrs. or min.
OCCUPATION (a) Trade, profession, or particular kind of work Mill Operative
(b) General nature of industry, business, or establishment in which employed (or employer) Cotton Mill
EDUCATIONAL ATTAINMENTS Common School
BIRTHPLACE S.C.
PARENTS
NAME OF FATHER Robert Moore
BIRTHPLACE OF FATHER (State or Country) Scm.
MAIDEN NAME OF MOTHER Agnes Campbell
BIRTHPLACE OF MOTHER (State or Country) SC
THE ABOVE IS TRUE TO THE BEST OF MY KNOWLEDGE
(Informant) Mrs Bonnie Moore
(Address) Gaffney SC
Filed 7-19, 1915 F. O. Hawley Registrar

MEDICAL CERTIFICATE OF DEATH
DATE OF DEATH July 16, 1915 (Month) (Day) (Year)
I HEREBY CERTIFY, That I attended deceased from , 19 , to , 19
that I last saw h alive on , 19
and that death occurred on the date above stated, at m.
The CAUSE OF DEATH* was as follows:
Fracture of Spinal Column
(185)
(Duration) yrs. mos. ds.
Contributory (Secondary) Paralysis
(Duration) yrs. mos. ds.
(Signed) J. R. Alexander, M.D.
July 16, 1915 (Address) Charlotte
*State the DISEASE CAUSING DEATH, or, in deaths from VIOLENT CAUSES, state (1) MEANS OF INJURY; and (2) whether ACCIDENTAL, SUICIDAL, or HOMICIDAL.
LENGTH OF RESIDENCE (For Hospitals, Institutions, Transients or Recent Residents)
At place of death yrs. mos. ds. In the State yrs. mos. ds.
Where was disease contracted, if not at place of death?
Former or usual residence.
PLACE OF BURIAL OR REMOVAL Gaffney SC
DATE OF BURIAL July 17, 1915
UNDERTAKER Zus Morrison
ADDRESS Charlotte

N. B.—Every item of information should be carefully supplied. AGE should be stated EXACTLY. PHYSICIANS should state CAUSE OF DEATH in plain terms, so that it may be properly classified. Exact statement of OCCUPATION is very important. See instructions on back of certificate.

Figure 1: Example of a North Carolina death certificate from 1915. Note that complete names and birthplaces of parents are given in the lower left portion of the certificate. Occupation, industry and educational attainment are also all given on the left side of the certificate.

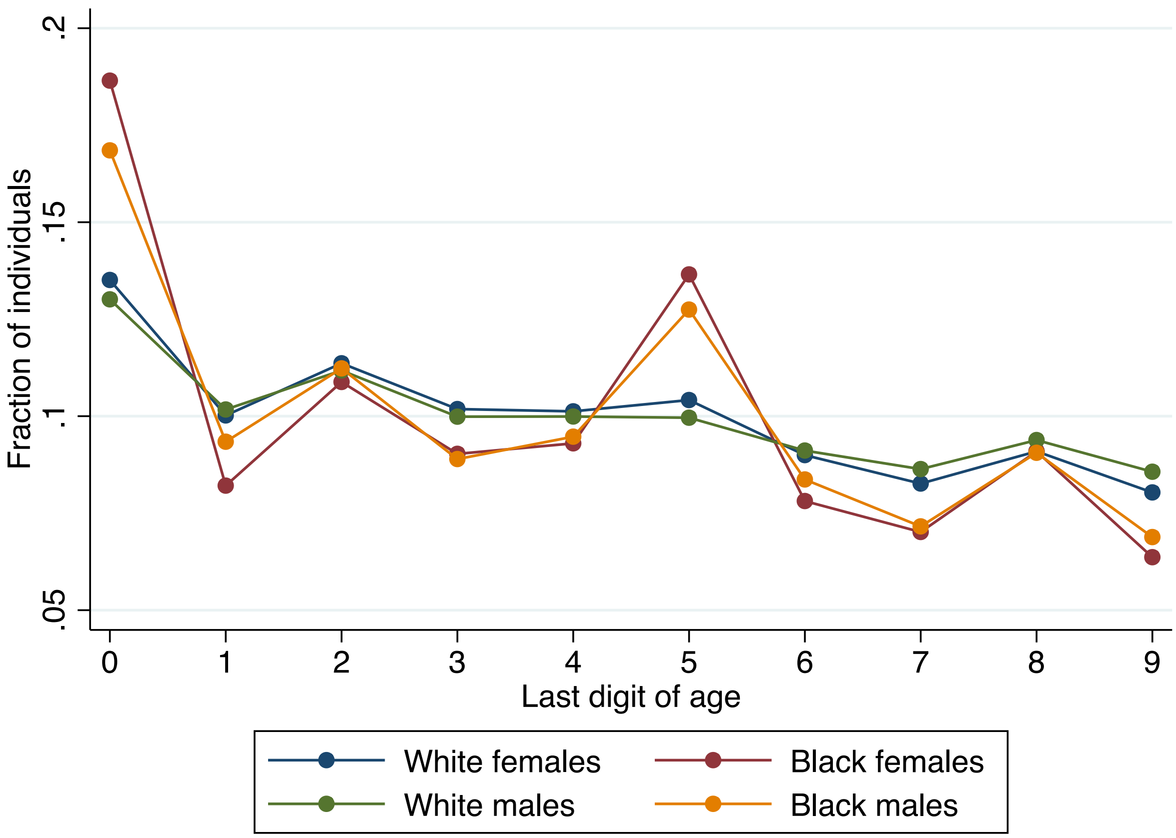


Figure 2: Distribution of last digit of age for individuals over the age of 19 by sex and race. Sample includes all individuals in North Carolina in the 1% IPUMS samples of the 1880, 1900, 1910, 1920 and 1930 federal censuses.

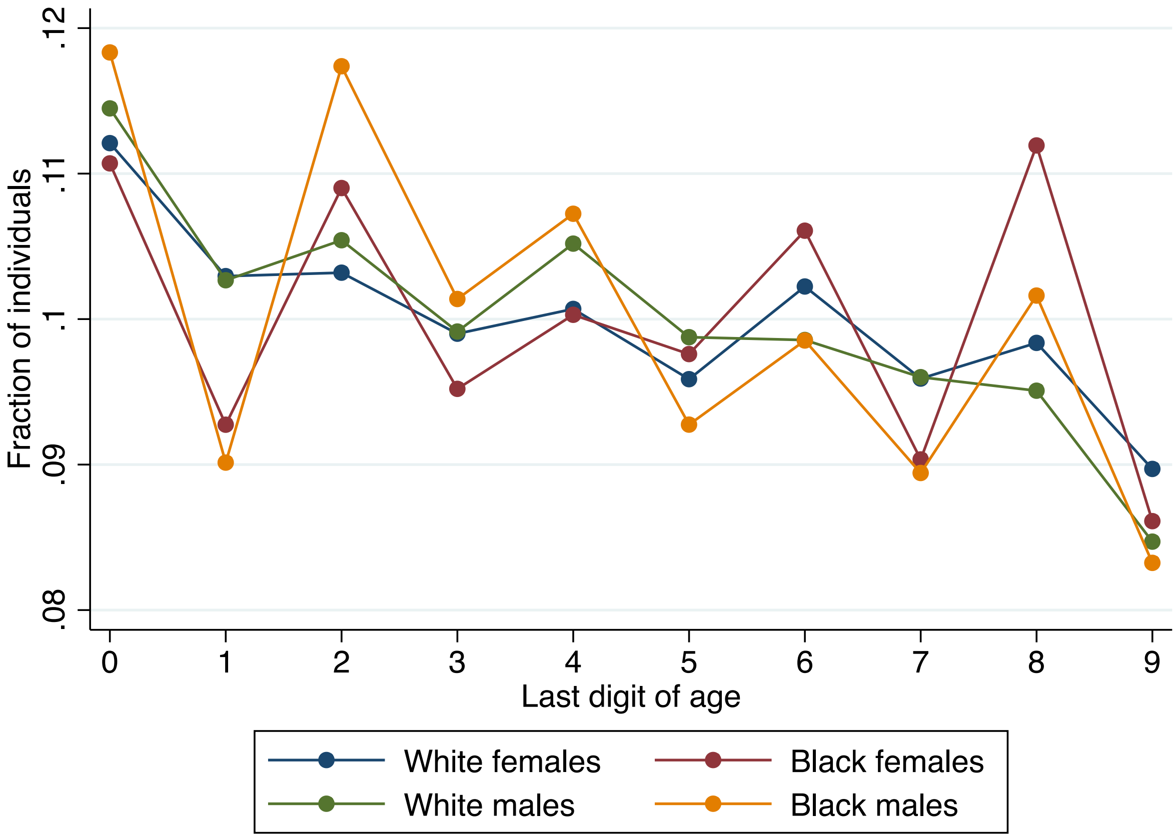


Figure 3: Distribution of last digit of age reported in the federal census for individuals over the age of 9 and under the age of 20 by sex and race. Sample includes all individuals in North Carolina in the 1% IPUMS samples of the 1880, 1900, 1910, 1920 and 1930 federal censuses.

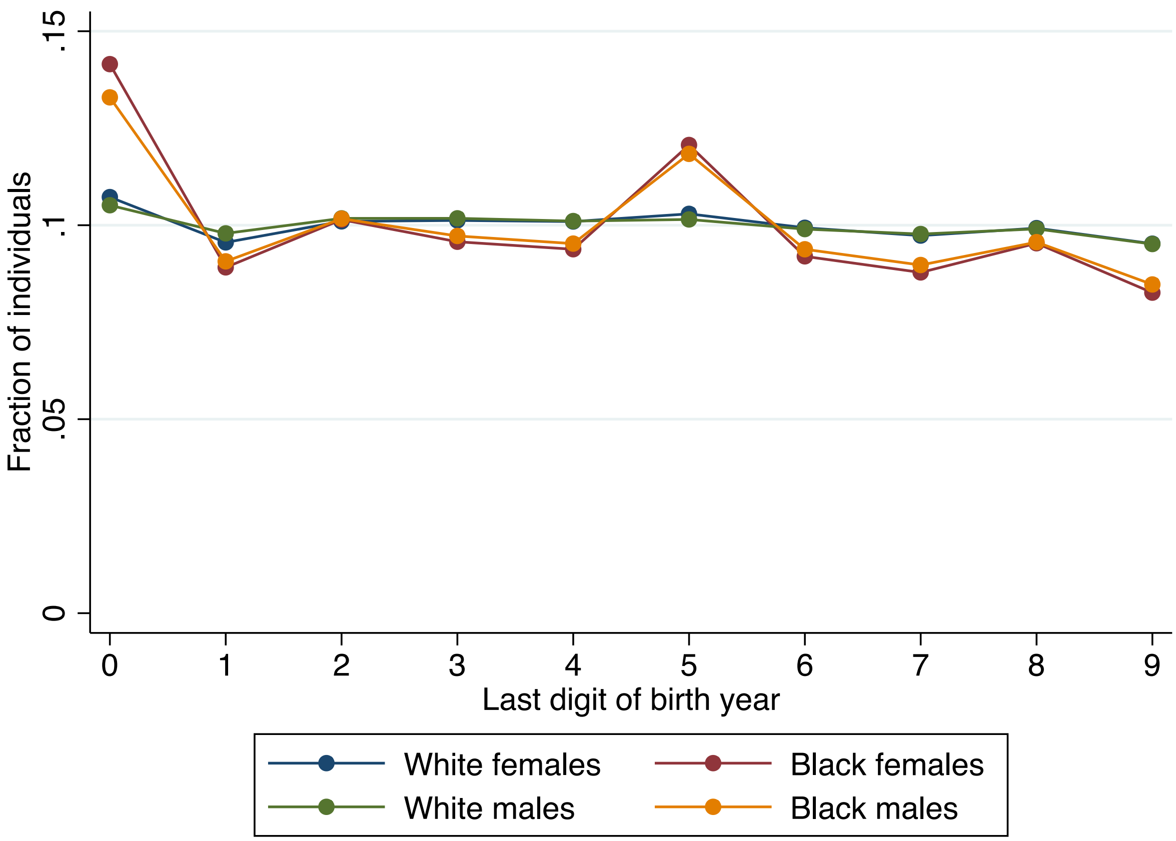


Figure 4: Distribution of last digit of birth year given in the death certificate for individuals over the age of 19 by sex and race. Sample includes all individuals in the universe of North Carolina death certificates from 1909 to 1975.

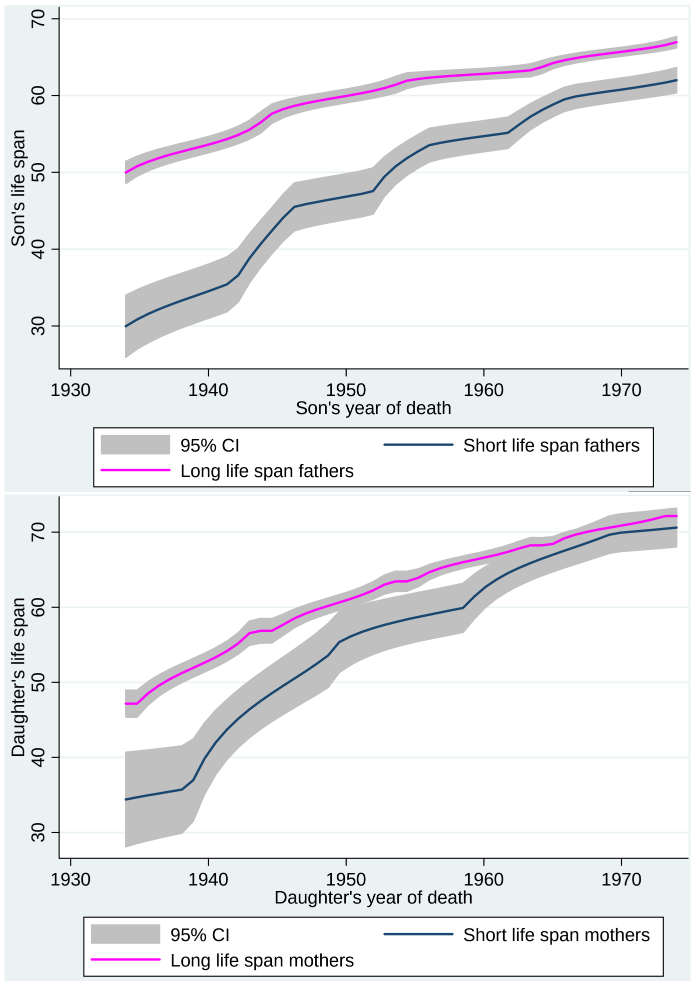


Figure 5: Longevity by cohort, gender and parental longevity. The upper panel is son's longevity across cohorts for father's with life spans shorter their predicted life span based on birth year and for father's with life spans longer than their predicted life span based on birth year. The lower panel is the equivalent graph for daughters based on their mother's life span. Both graphs are kernal-weighted local polynomial fits.

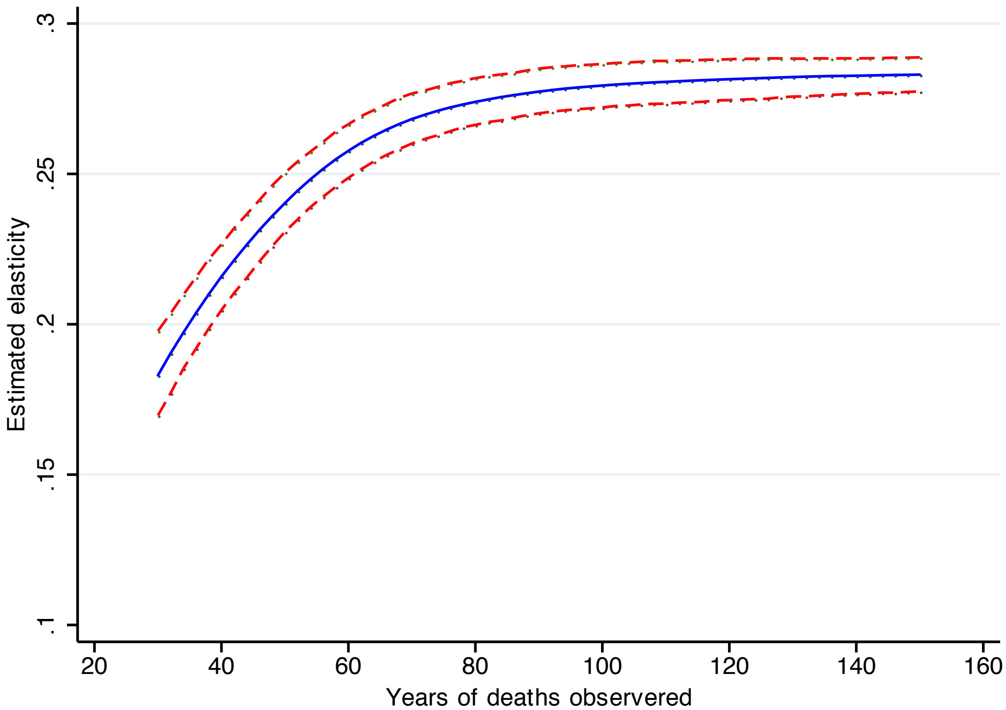


Figure 6: Estimated intergenerational longevity elasticity by level of truncation. The solid line gives the mean estimated value of the elasticity for each level of truncation. The dashed lines give the 5th and 95th percentiles of the distribution of estimated values at each level of truncation. The true value of the elasticity used in the simulations is 0.28.

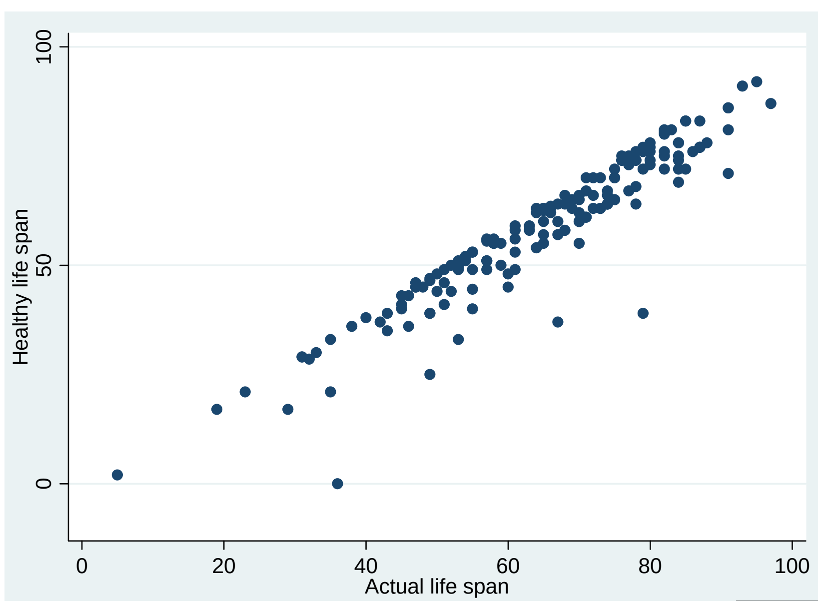


Figure 7: Healthy life span versus actual life span. Healthy life span is calculated as actual life span minus years spent in poor health where years in poor health come from the durations of significant medical conditions reported on the death certificate.

Table 1: Linking success rates for a 10% sample of individuals from Mecklenburg County dying between 1934 and 1974 by age and race.

	All individuals	Males	Females	White	Black
Number of individuals in initial sample	12,317	5,766	4,604	7,083	3,283
Number of individuals born before 1930	10,104	4,625	3,889	6,000	2,511
Linking child to federal census					
% of original sample not found in census	68.28%	68.54%	68.71%	61.44%	83.88%
% of original sample found but not living with parents	0.14	0.12	0.19	0.22	--
% of original sample found living with mother but not father	0.48	0.43	0.21	0.42	0.15
% of original sample found living with father but not mother	0.27	--	0.06	0.01	0.06
% of original sample found living with both parents	30.84	30.91	30.82	37.9	15.91
Linking parents to death certificates					
% of original sample linked to father's death certificate only	4.85	5.04	4.55	6.13	2.04
% of original sample linked to mother's death certificate only	4.67	4.71	4.59	5.81	2.22
% of original sample linked to both parents' death certificates	5.23	5.09	4.66	6.64	1.19
Number of individuals matched to at least one parent's death certificate	1521	872	649	1338	184

Table 2: Summary statistics by linking outcome, gender and race.

	Sons		Daughters		White individuals		Black individuals	
	Not matched to either parent	Matched to at least one parent	Not matched to either parent	Matched to at least one parent	Not matched to either parent	Matched to at least one parent	Not matched to either parent	Matched to at least one parent
Age at death	60.8	61.2	63.8	65.0	66.0	64.2	54.7	52.9
	(17.5)	(15.6)	(19.3)	(16.9)	(17.3)	(15.6)	(18.3)	(17.6)
% born in North Carolina	56.1%	84.2%	55.8%	79.0%	57.6%	83.9%	53.0%	73.7%
% born in South Carolina	22.3	8.1	23.4	12.1	15.0	7.5	37.2	26.3
% in farming	--	18.5	--	0.6	--	13.6	--	2.2
% in skilled/semi-skilled occ	--	32.2	--	32.6	--	33.9	--	15.2
% in textiles	--	14.6	--	10.3	--	14.0	--	4.2
% in unskilled occupations	--	8.5	--	32.3	--	10.6	--	78.3
% in white collar occupatioı	--	40.9	--	34.3	--	41.9	--	4.4
% male	100.0	100.0	0.0	0.0	44.1	58.6	54.0	50.3
% white	65.5	89.6	63.8	86.1	100.0	100.0	0.0	0.0
Number of individuals	5,006	867	4,066	638	5,874	1,326	3,196	179

Notes: Standard deviation for age at death is given in parentheses. Occupations have only been coded for individuals with at least one parent linked to the death certificates. The occupational distributions are percentages relative to all individuals reporting an occupation. Individuals with textile occupations are also counted in their relevant skill category (skilled, unskilled, white collar).

Table 3: Average age misreporting by gender and race.

	Death certificate birth year - census birth year			Absolute value of (death certificate birth year - census birth year)		
	Child	Mother	Father	Child	Mother	Father
All	-0.29	-0.58	-0.44	0.98	1.61	1.55
	(1.58)	(2.45)	(2.45)	(1.27)	(1.93)	(1.95)
Male	-0.39	-0.57	-0.53	0.96	1.66	1.54
	(1.47)	(2.46)	(2.42)	(1.17)	(1.89)	(1.94)
Female	-0.21	-0.59	-0.38	0.98	1.59	1.53
	(1.62)	(2.48)	(2.43)	(1.31)	(1.99)	(1.93)
White	-0.47	-0.79	-0.57	0.83	1.41	1.42
	(1.20)	(2.05)	(2.20)	(.99)	(1.68)	(1.77)
Black	0.56	1.08	0.48	1.73	3.39	2.68
	(2.55)	(4.21)	(3.85)	(1.95)	(2.71)	(2.94)

Notes: Census birth year is calculated by subtracting the age reported in the census from the year of the census. This imputed birth year may be one year off of the death certificate birth year simply because the individual has not reached her birthday by the time of the census.

Table 4: Age heaping for death certificates and the federal census by gender and race.

	Percentage of observations with a last digit of 0 or 5					
	Child's census age	Father's census age	Mother's census age	Child's death certificate birth year	Father's death certificate birth year	Mother's death certificate birth year
All	18.7%	24.4%	23.7%	20.4%	18.6%	19.7%
Male	17.8	24.3	23.0	19.3	17.3	20.6
Female	19.9	24.5	24.6	18.5	20.6	18.6
White	18.9	23.4	22.9	19.5	18.7	19.7
Black	17.6	29.3	27.8	17.9	18.5	20.4

Table 5: OLS estimates of age heaping correlates for children.

Age range:	10 to 19 years old			
	Age ends in zero or five		Age ends in an even number	
Dependent variable:	(1=yes)		(1=yes)	
Black (1=yes)	0.0049	0.0293	0.0181***	0.0324
	(0.0028)	(0.0293)	(0.0040)	(0.0243)
Male (1=yes)	0.0056*	0.0060	0.0064	0.0064
	(0.0028)	(0.0032)	(0.0034)	(0.0041)
Male x Black		-0.0012		-0.0001
		(0.0050)		(0.0093)
Lives on a farm (1=yes)	0.0043	0.0062	0.0041	-0.0003
	(0.0035)	(0.0040)	(0.0031)	(0.0059)
Lives on farm x Black		-0.0023		0.0150*
		(0.0046)		(0.0068)
Household head's log income	0.0053*	0.0087	-0.0041	-0.0015
	(0.0021)	(0.0050)	(0.0129)	(0.0128)
Household head's log income x Black		-0.0079		-0.0152**
		(0.0114)		(0.0039)
Household head's literacy (1=literate)	0.0028	0.0018	-0.0030	-0.0039
	(0.0030)	(0.0050)	(0.0072)	(0.0144)
Household head's literacy x Black		0.0022		0.0028
		(0.0087)		(0.0161)
Year indicators:				
1880	-0.0166***	-0.0163***	-0.0330***	-0.0327***
	(0.0004)	(0.0005)	(0.0011)	(0.0013)
1900	-0.0294***	-0.0292***	-0.0599***	-0.0603***
	(0.0030)	(0.0007)	(0.0024)	(0.0019)
1910	-0.0278***	-0.0277***	-0.0348***	-0.0356***
	(0.0007)	(0.0007)	(0.0032)	(0.0025)
1920	-0.0212***	-0.0211***	-0.0580***	-0.0591***
	(0.0009)	(0.0011)	(0.0041)	(0.0031)
1930	-0.0297***	-0.0297***	-0.0562***	-0.0577***
	(0.0011)	(0.0013)	(0.0047)	(0.0036)
Constant	0.2166***	0.2043***	0.5781***	0.5795***
	(0.0098)	(0.0177)	(0.0437)	(0.0555)
Observations	67,693	67,693	67,693	67,639

Robust standard errors clustered by census year are included in parentheses. The dependent variable equals one if the individual's age ends in zero or five and equals zero otherwise. Regression sample includes all North Carolina residents in the 1% IPUMS samples of the 1870, 1880, 1900, 1910, 1920 and 1930 federal censuses. The omitted category for the census year indicators is 1870. * significant at 10%, ** significant at 5%, *** significant at 1%

Table 6: OLS estimates of age heaping correlates for adults, age ending in zero or five (1=yes) as dependent variable.

Age range:	20 to 59 years old	
Black (1=yes)	0.0549***	0.0659**
	(0.0067)	(0.0200)
Male (1=yes)	-0.0115**	-0.0053
	(0.0040)	(0.0031)
Male x Black		-0.0218***
		(0.0037)
Lives on a farm (1=yes)	-0.0126*	-0.0072***
	(0.0051)	(0.0014)
Lives on farm x Black		-0.0088
		(0.0115)
Household head's log income	-0.0115***	-0.0054
	(0.0021)	(0.0034)
Household head's log income x Black		-0.0169***
		(0.0015)
Own literacy (1=literate)	-0.0439***	-0.0254**
	(0.0062)	(0.0089)
Own literacy x Black		-0.0336***
		(0.0063)
Age	-0.0020***	-0.0027***
	(0.0002)	(0.0001)
Age x Black		0.0024***
		(0.0002)
Year indicators:		
1880	-0.0425***	-0.0422***
	(0.0005)	(0.0006)
1900	-0.0830***	-0.0821***
	(0.0015)	(0.0019)
1910	-0.0968***	-0.0957***
	(0.0019)	(0.0025)
1920	-0.0969***	-0.0959***
	(0.0022)	(0.0030)
1930	-0.0882***	-0.0876***
	(0.0020)	(0.0028)
Constant	0.4828***	0.4618***
	(0.0082)	(0.0076)
Observations	119,412	119,412

Robust standard errors clustered by census year are included in parentheses. The dependent variable equals one if the individual's age ends in zero or five and equals zero otherwise. Regression sample includes all North Carolina residents in the 1% IPUMS samples of the 1870, 1880, 1900, 1910, 1920 and 1930 federal censuses. The omitted category for the census year indicators is 1870. * significant at 10%, ** significant at 5%, *** significant at 1%

Table 7: Occupational distribution by birth cohort.

	Distribution of ocupations within cohort			
		Skilled, semi-		
Decade of birth	Farmer	skilled	Unskilled	White collar
1830s	67%	17%	0%	17%
1840s	75	16	4	5
1850s	68	11	3	19
1860s	69	11	8	12
1870s	57	16	8	18
1880s	44	22	10	24
1890s	30	27	11	32
1900s	19	24	11	45
1910s	10	45	10	34
1920s	5	35	10	50

Note: Distributions are based on all males in the sample including both fathers and male children.

Table 8: Occupational transitions.

		Father's occupation				
			Skilled, semi-			
		Farmer	skilled	Unskilled	White collar	Total
Son's Occupation	Farmer	50	4	1	0	55
	Skilled, semi-skilled	45	15	7	10	77
	Unskilled	13	3	6	1	23
	White collar	54	16	6	35	111
	Total	162	38	20	46	266

Table 9: Mean life span by occupational category.

Occupational category	Son's life span by son's occupation	Son's life span by father's occupation	Daughter's life span by father's occupation
Farmer	68.2	64.7	67.6
	(15.0)	(13.9)	(16.6)
Skill, semi-skilled	62.0	56.8	64.1
	(13.5)	(16.9)	(19.2)
Unskilled	51.7	53.2	55.4
	(17.8)	(18.1)	(17.5)
White collar	60.4	58.4	61.5
	(12.2)	(15.3)	(16.4)
Number of observations	360	414	316

Notes: Standard deviations are given in parentheses. Life span is calculated using the year of death and year of birth from the death certificate.

Table 10: Effects of occupation and occupational transitions on change in longevity, son's life span minus father's life span as dependent variable.

	Sons	Sons	Sons	Daughters
Son's occupation dummies				
Skilled/semi-skilled	0.11***	--	0.11***	--
	(0.03)	--	(0.03)	--
White collar	0.07***	--	0.08***	--
	(0.02)	--	(0.03)	--
Farmer	0.08**	--	0.07*	--
	(0.03)	--	(0.03)	--
Father's occupation dummies				
Skilled/semi-skilled	--	0.01	0.01	-0.06
	--	(0.04)	(0.04)	(0.06)
White collar	--	0.03	0.01	-0.001
	--	(0.04)	(0.04)	(0.040)
Farmer	--	0.07***	0.07***	0.001
	--	(0.03)	(0.02)	(0.035)
Child's birth year	0.84***	0.96***	0.76***	1.07***
	(0.20)	(0.27)	(0.22)	(0.38)
(Child's birth year)^2	-0.0002***	-0.0003***	-0.0002***	-0.0003***
	(0.0001)	(0.0001)	(0.0001)	(0.0001)
Constant	-783.64***	-897.16	-710.24***	-955.08***
	(191.58)	(253.97)	(206.25)	(357.91)
Observations	578	562	545	411
R-squared	0.36	0.36	0.37	0.39

Robust standard errors in parentheses. * significant at 10%; ** significant at 5%; *** significant at 1%. Ommitted occupational dummy is unskilled for both son's and father's occupation. Child's life span is calculated as the death year given on the death certificate minus the birth year given on the death certificate.

Table 11: Intergenerational longevity elasticity regressions, log of child's life span as dependent variable.

	Sons				Daughters			
Model:	OLS	2SLS IV	OLS	2SLS IV	OLS	2SLS IV	OLS	2SLS IV
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
ln(Parent's life span)	0.28***	0.28***	0.20***	0.21***	0.19**	0.23***	0.19***	0.26***
	(0.09)	(0.07)	(0.07)	(0.08)	(0.08)	(0.06)	(0.07)	(0.07)
ln(Parent's life span) x			0.12	0.10*			-0.01	-0.06
(child's birth year-1900)/10			(0.09)	(0.06)			(0.07)	(0.06)
Child's birth year	1.00***	1.00***	0.82***	0.85***	0.79***	0.79***	0.81***	0.90***
	(0.24)	(0.18)	(0.22)	(0.20)	(0.20)	(0.17)	(0.25)	(0.21)
(Child's birth year)^2	-0.0003***	-0.0003***	-0.0002***	-0.0002***	-0.0002***	-0.0002***	-0.0002***	-0.0002***
	(0.0001)	(0.00005)	(0.0001)	(0.0001)	(0.0001)	(0.00005)	(0.0001)	(0.0001)
Constant	-938.74***	-938.70***	-715.76***	-753.53***	-736.20***	-733.23***	-758.85***	-856.60***
	(227.46)	(173.92)	(215.93)	(203.33)	(189.50)	(167.23)	(250.99)	(210.27)
Observations	585	585	585	585	424	424	424	424
R-squared	0.37	0.38	0.38	0.38	0.41	0.41	0.41	0.41

Notes: Robust standard errors in parentheses. * significant at 10%, ** significant at 5%, *** significant at 1%. Father's life span is used as the parental life span in the regressions for sons; mother's life span is used in the regressions for daughters. For the OLS regresions, all life spans are calculated by subtracting year of death on the death certificate from year of birth given on the death certificate. In the two-stage least squares regressions, parental life spans based on the birth year given in the census are used to instrument for life spans based on the birth year given on the death certificate.

Table 12: Intergenerational longevity elasticity regressions including both parents, log of child's life span as dependent variable.

	Sons		Daughters	
	(1)	(2)	(3)	(4)
ln(Father's life span)	0.36***	0.16*	0.09	0.07
	(0.12)	(0.09)	(0.12)	(0.10)
ln(Father's life span) x		0.28**		0.06
(child's birth year-1900)/10		(0.13)		(0.14)
ln(Mother's life span)	0.16*	0.07	0.32***	0.27**
	(0.09)	(0.08)	(0.12)	(0.10)
ln(Mother's life span) x		0.09		0.09
(child's birth year-1900)/10		(0.08)		(0.10)
Child's birth year	1.52***	0.95***	0.81**	0.63*
	(0.45)	(0.29)	(0.33)	(0.33)
(Child's birth year)^2	-0.0004***	-0.0003***	-0.0002**	-0.0002**
	(0.0001)	(0.0001)	(0.0001)	(0.0001)
Constant	-1427.45***	-739.40*	-749.34**	-521.20
	(424.01)	(287.72)	(316.99)	(327.52)
Observations	293	293	215	215
R-squared	0.41	0.44	0.41	0.41

Notes: Robust standard errors in parentheses. * significant at 10%, ** signficant at 5%, *** significant at 1%. Life span is defined as death certificate death year minus death certificate birth year.

Table 13: Correlations in spouses' age at death using parents in the linked sample.

	Fathers		Mothers	
Dependent variable:	Age at death	ln(Age at death)	Age at death	ln(Age at death)
Spouse's age at death	0.142***		0.179***	
	(0.038)		(0.047)	
ln(Spouse's age at death)		0.162***		0.196***
		(0.039)		(0.052)
Own birth year - 1870	-0.169***	-0.003***	-0.169***	-0.002***
	(0.033)	(0.001)	(0.035)	(0.000)
(Own birth year - 1870)^2	0.002	0.000	-0.008	-0.000**
	(0.002)	(0.000)	(0.003)	(0.000)
Constant	61.368***	3.567***	63.131***	3.474***
	(2.849)	(0.167)	(3.591)	(0.222)
Observations	620	619	620	619

Robust standard errors given in parentheses. Sample is restricted to white families. Life span is defined as the death year given on the death certificate minus the birth year given on the death certificate. * significant at 10%, ** significant at 5%, *** significant at 1%

Table 14: Modern intergenerational correlations in longevity using IHIS data.

	White males						White females					
Dependent variable:	Age at death			Ln(Age at death)			Age at death			Ln(Age at death)		
Father's age at death	0.011		0.022				0.023**		0.027*			
	(0.018)		(0.011)				(0.007)		(0.009)			
Mother's age at death		0.012	0.010					0.023**	0.020*			
		(0.016)	(0.014)					(0.007)	(0.007)			
ln(Father's age at death)				0.002		0.013				0.015*		0.018*
				(0.015)		(0.008)				(0.006)		(0.007)
ln(Mother's age at death)					0.007	0.007					0.015*	0.012
					(0.012)	(0.011)					(0.006)	(0.006)
Number of observations	1806	1577	1433	1806	1577	1433	2561	2439	2209	2561	2439	2209

Robust standard errors clustered by region are included in parentheses. All regressions control for a quadratic in the child's birth year and region fixed effects. * significant at 10%, ** significant at 5%, *** significant at 1%

Table 15: Intergenerational longevity elasticities restricting age at death using IHIS data.

	White males			White females		
Minimum age cutoffs applied to:	Son	Father	Both	Daughter	Mother	Both
Minimum age at death cutoff:						
None	0.002	0.002	0.002	0.015*	0.015*	0.015*
	(0.015)	(0.015)	(0.015)	(0.006)	(0.006)	(0.006)
	N=1806	N=1806	N=1806	N=2439	N=2439	N=2439
30 years old	0.002	0.002	0.002	0.015*	0.017**	0.017**
	(0.015)	(0.017)	(0.017)	(0.006)	(0.004)	(0.004)
	N=1805	N=1786	N=1785	N=2439	N=2380	N=2380
40 years old	0.004	0.006	0.008	0.015*	0.028***	0.027***
	(0.0142)	(0.016)	(0.016)	(0.006)	(0.004)	(0.004)
	N=1798	N=1729	N=1721	N=2439	N=2276	N=2276
50 years old	0.005	0.033	0.033	0.014*	0.035***	0.035***
	(0.010)	(0.029)	(0.024)	(0.005)	(0.006)	(0.005)
	N=1762	N=1605	N=1571	N=2426	N=2121	N=2114
60 years old	0.007	0.052	0.067	0.015**	0.029**	0.026*
	(0.010)	(0.043)	(0.055)	(0.004)	(0.009)	(0.010)
	N=1645	N=1384	N=1299	N=2374	N=1895	N=1856

Robust standard errors clustered by region are given in parentheses. Sample sizes are given below the standard errors. The reported result is the coefficient on the natural log parent's age at death in a regression that controls for a quadratic in the child's year of birth and include region fixed effects. * significant at 10%, ** significant at 5%, *** significant at 1%

Table 16: Correlations in spouses' age at death using IHIS data.

Dependent variable:	Age at death	Ln(Age at death)	Age at death	Ln(Age at death)
	White fathers		White mothers	
Spouse's age at death	0.103***		0.125***	
	(0.010)		(0.010)	
ln(Spouse's age at death)		0.100***		0.132***
		(0.010)		(0.008)
Number of observations	5296	5296	5296	5296
	Black fathers		Black mothers	
Spouse's age at death	0.146*		0.198*	
	(0.055)		(0.074)	
ln(Spouse's age at death)		0.143**		0.245**
		(0.040)		(0.064)
Number of observations	624	624	624	624

Robust standard errors clustered by region are given in parentheses. Race of parent's is determined by the reported race of the child. All regressions control for a quadratic in the child's year of birth and include region fixed effects. * significant at 10%, ** significant at 5%, *** significant at 1%

A Death Certificate Variables

The format of the North Carolina death certificates has changed over time. Below are descriptions of the information provided on the death certificate in five year intervals. The information on sex, race, date of birth, place of birth, date of death, death location, name of father, name of mother, name of spouse and residence are all indexed electronically and searchable. The remaining variables must be read directly from images of the original death certificate.

1910: Sex, birth, race, date of birth, date of death, marital status, name of father, birthplace of father, place of birth, maiden name of mother, birthplace of mother, occupation, place of burial, cause of death, contributory factors, length of hospital stay (if applicable)

1915: Sex, birth, race, date of birth, date of death, marital status, name of father, birthplace of father, place of birth, maiden name of mother, birthplace of mother, occupation, place of burial, cause of death, contributory factors, length of hospital stay (if applicable), educational attainment

1920: Sex, birth, race, date of birth, date of death, marital status, name of father, birthplace of father, place of birth, maiden name of mother, birthplace of mother, occupation, general industry, name of employer, place of burial, cause of death, contributory factors, length of hospital stay (if applicable), educational attainment, where disease was contracted, was test done to confirm diagnosis, did operation precede death, was there autopsy

1925: Sex, birth, race, date of birth, date of death, marital status, name of father, birthplace of father, place of birth, maiden name of mother, birthplace of mother, occupation, general industry, name of employer, place of burial, cause of death, contributory factors, length of hospital stay (if applicable), educational attainment, where disease was contracted, was test done to confirm diagnosis, did operation precede death, was there autopsy

1930: Sex, birth, race, date of birth, date of death, marital status, name of father, birthplace of father, place of birth, maiden name of mother, birthplace of mother, occupation, general industry, name of employer, place of burial, cause of death, contributory factors, length of hospital stay (if applicable), educational attainment, where disease was contracted, was test done to confirm diagnosis, did operation precede death, was there autopsy

1935: Sex, birth, race, date of birth, date of death, marital status, name of father, birthplace of father, place of birth, maiden name of mother, birthplace of mother, occupation, general industry, name of employer, date last worked occupation, how long in occupation, place of burial, cause of death, contributory factors, length of hospital stay (if applicable), educational attainment, where disease was contracted, was test done to confirm diagnosis, did operation precede death, was there autopsy, was death accident/suicide/homicide, where injury occurred, nature of injury

1940: Sex, birth, race, date of birth, date of death, marital status, name of father, birthplace of father, place of birth, maiden name of mother, birthplace of mother, occupation, general industry, name of employer, date last worked occupation, how long in occupation, place of burial, cause of death, contributory factors, length of hospital stay (if applicable), educational attainment, where disease was contracted, was test done to confirm diagnosis, did operation precede death, was there autopsy, was death accident/suicide/homicide, where injury occurred, nature of injury

1945: Sex, birth, race, date of birth, date of death, marital status, name of father, birthplace of father, place of birth, maiden name of mother, birthplace of mother, occupation, general industry, name of employer, date last worked occupation, how long in occupation, place of burial, cause of death, interval between onset and death, contributory factors, other conditions, ever in armed forces, social security number, length of hospital stay (if applicable), educational attainment, where disease was contracted, was test done to confirm diagnosis, did operation precede death, was there autopsy, was death accident/suicide/homicide, where injury occurred, nature of injury

1955: Sex, birth, race, date of birth, date of death, marital status, name of father, birthplace of father, place of birth, maiden name of mother, birthplace of mother, occupation, general industry, name of employer, date last worked occupation, how long in occupation, place of burial, cause of death, interval between onset and death, contributory factors, other conditions, ever in armed forces, social security number, length of hospital stay (if applicable), educational attainment, where disease was contracted, was test done to confirm diagnosis, did operation precede death, was there autopsy, was death accident/suicide/homicide, where injury occurred, nature of injury

1960: Sex, birth, race, date of birth, date of death, marital status, name of father, birthplace of father, place of birth, maiden name of mother, birthplace of mother, occupation, general industry, name of employer, date last worked occupation, how long in occupation, place of burial, cause of death, interval between onset and death, contributory factors, other conditions, ever in armed forces, social security number, length of hospital stay (if applicable), educational attainment, where disease was contracted, was test done to confirm diagnosis, did operation precede death, was there autopsy, was death accident/suicide/homicide, where injury occurred, nature of injury

1965: Sex, birth, race, date of birth, date of death, marital status, name of father, birthplace of father, place of birth, maiden name of mother, birthplace of mother, occupation, general industry, name of employer, date last worked occupation, how long in occupation, place of burial, cause of death, interval between onset and death, contributory factors, other conditions, ever in armed forces, social security number, length of hospital stay (if applicable), educational attainment, where disease was contracted, was test done to confirm diagnosis, did operation precede death, was there autopsy, was death accident/suicide/homicide, where injury occurred, nature of injury

1970: Sex, birth, race, date of birth, date of death, marital status, name of father, birthplace of father, place of birth, maiden name of mother, birthplace of mother, occupation, general industry, name of employer, date last worked occupation, how long in occupation, place of burial, cause of death, interval between onset and death, contributory factors, other conditions, ever in armed forces, social security number, length of hospital stay (if applicable), educational attainment, where disease was contracted, was test done to confirm diagnosis, did operation precede death, was there autopsy, was death accident/suicide/homicide, where injury occurred, nature of injury, physician certification

1975: Sex, birth, race, date of birth, date of death, marital status, name of father, birthplace of father, place of birth, maiden name of mother, birthplace of mother, occupation, general industry, name of employer, date last worked occupation, how long in occupation, place of burial, cause of death, interval between onset and death, contributory factors, other conditions, ever in armed forces, social security number, length of hospital stay (if applicable), educational attainment, where disease was contracted, was test done to confirm diagnosis, did operation precede death, was there autopsy, was death accident/suicide/homicide, where injury occurred, nature of injury, physician certification