



**Gender and Intergenerational Mobility: Using Health
Outcomes to Compare Intergenerational Mobility Across
Gender and Over Time***

John Parman

The College of William and Mary

College of William and Mary
Department of Economics
Working Paper Number 122

April 2012

* 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. This is a very preliminary draft, please do not cite without author's permission. For the most recent version of the paper, please see <http://wmpeople.wm.edu/jmparman>.

COLLEGE OF WILLIAM AND MARY
DEPARTMENT OF ECONOMICS
WORKING PAPER # 122
April 2012

Gender and Intergenerational Mobility: Using Health Outcomes to Compare Intergenerational Mobility Across Gender and Over Time

Abstract

Changes in intergenerational mobility over time have been the focus of extensive research. However, existing studies have been limited to studying only males and relying on intergenerational correlations in outcome variables that often lack clear welfare implications. This paper introduces a new methodology for measuring 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. It introduces a new intergenerational dataset spanning seven decades that is constructed by linking individuals' death certificates to those of their parents. Relying on death certificates data allows for linking both males and females to their parents. Life span calculated from these death certificates provides a measure of welfare that has a consistent interpretation across time and genders. Intergenerational correlations in life span serve as our measure of mobility. We 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.

John Parman
Department of Economics,
The College of William and Mary
Williamsburg, USA
jmparman@wm.edu

Gender and Intergenerational Mobility: Using Health Outcomes to Compare Intergenerational Mobility Across Gender and Over Time

John Parman*

April 27, 2012

Abstract

Changes in intergenerational mobility over time have been the focus of extensive research. However, existing studies have been limited to studying only males and relying on intergenerational correlations in outcome variables that often lack clear welfare implications. This paper introduces a new methodology for measuring 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. It introduces a new intergenerational dataset spanning seven decades that is constructed by linking individuals' death certificates to those of their parents. Relying on death certificate data allows for linking both males and females to their parents. Life span calculated from these death certificates provides a measure of welfare that has a consistent interpretation across time and genders. Intergenerational correlations in life span serve as our measure of mobility. We 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 is how the economic gains experienced by the economy as a whole have been 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

*jparman@wm.edu, Department of Economics, College of William & Mary; 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. This is a very preliminary draft, please do not cite without author's permission. For the most recent version of the paper, please see <http://wmpeople.wm.edu/jparman>.

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.

This paper takes a new approach to measuring intergenerational mobility, one that circumvents several of the limitations of prior mobility measures. We opt for measuring mobility by the strength of the relationship between long term health outcomes of children and their parents, specifically by estimating the strength of the correlation between a child's life span and that of his or her parents. This approach addresses two major constraints faced by the existing literature on historical mobility rates. First, our estimates rely on a variable that has a clear interpretation in terms of welfare. An increase in longevity is a clearer indication of an improvement in welfare compared to transitions between broad occupational categories typically used to measure mobility. 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 eighteenth century or the nineteenth 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 farmer in nineteenth century 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, income and 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 the female population. Focusing on life span produces measures of mobility that are equally meaningful for males and females. The nature of the death certificates from which life span is drawn makes it possible to match adult females to their parents, something 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 represents a major improvement in our understanding of historical mobility patterns.

Our intergenerational health data present evidence that is largely complementary to existing measures of historical mobility rates. We find that there is a strong correlation between an

individual's longevity and that of their parents, both for males and females. This correlation has been growing stronger for males over the past century but has remained relatively stable over time for females. 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. The results of this paper suggest that longevity and other long term health outcomes offer a promising new direction for the estimation of historical mobility patterns.

The remainder of this paper is organized as follows. The next section reviews the existing literature on historical mobility patterns, highlighting the significant stylized facts that have emerged and discussing the difficulties in interpreting these stylized facts. Section 3 describes our approach of using health measures to estimate intergenerational mobility. Section 4 discusses the construction of an intergenerational dataset linking health and occupational outcomes of children to those of their parents. Section 5 presents estimates of the intergenerational correlation in health outcomes across gender and over time. The final sections discuss directions for future research using these health data to further explore 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 for several developed countries including the United States. These longitudinal studies allow for comparing the incomes and occupations of sons and daughters to their parents. The 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 and Mazumder (2005) for more recent evidence that these may be underestimates of the true elasticity). 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).

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

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 across genders. It is unclear how one should interpret observed earnings if work decisions are not based solely on maximizing 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 be used for historical studies due to the lack of historical income data. The primary approach to measuring historical mobility rates has been to instead 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. However, the last publicly available census is the 1930 federal census. Consequently, estimating historical intergenerational income elasticities is not possible.

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 by introducing gender differences in mobility. With the labor force participation of females and the 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 the 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 given name on the census will be the same when he is a child as 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, an insufficient amount of information for uniquely identifying the

female as a child.

3 Using Health Outcomes to Estimate Mobility

This paper will take an alternative approach to estimating historical mobility rates that addresses several of these issues 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 several 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 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 studies, death certificates offer a larger sample of individuals to study that covers a longer time period: several states have maintained death certificates for their complete state populations back to the late nineteenth century. Relative to historical studies focused census data, 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 to instead 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). Once again, the experience of a group that is by definition a set of outliers is not a particularly useful way to assess the experience of the population.

An alternative approach in the literature on intergenerational correlations in health has been to choose health measures 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, we can construct measures of intergenerational correlations in health that use comparable health measures for parents and children that can be observed over the past 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 outcomes based on linked death certificate 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 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 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, not individual death certificates. 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.

Our 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 by name and date of birth 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 that 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.²

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 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 significant medical conditions to create a measure of healthy life span (or alternatively a measure of years spent in poor health). 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, we will focus on measures of longevity. As we expand the sample we will turn our attention to these other measures of health outcomes.

The information on age and parent's names makes it possible to link individuals to their parent's households 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 us to match these long term health outcomes of children to the exact same measures for their parents. The availability of the parent's names on the death certificate 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 our study. The details of this linking process between death certificates and censuses are describing in the following section.

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 us to search 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

²For a complete listing of the variables reported by year of death certificate, see the appendix.

County. Mecklenburg was chosen because it contains the city of Charlotte, giving us a far more heterogeneous sample in terms of rural and urban places of residence and occupational distribution than the other counties in the state. The other counties will be sampled in future waves of the data collection process.

To create a sample of individuals for a particular year, we 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. For the preliminary dataset used in this paper, we have generated a ten percent sample by sampling every tenth year beginning with 1934. The linking of the remaining years is currently underway.

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, we search for individuals by full name, birth year and birth state in the earliest federal census for which the person was alive (if an individual's birth year was during a census year, we search the next census). Choosing the earliest possible census affords the best chance to find the individual still living in her parents' household. We use the parents' names given on the death certificate 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. The variables include household size, household location, and the parents' ages, birthplaces and occupations.

Given the additional variables from the census, we now have sufficient information to uniquely match the parents to their death certificates. We 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, information from both the parent's death certificate and the child's death certificate is transcribed from images of the original certificates. This adds information on

occupation, 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 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 the 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) and 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. Roughly one fifth of the individuals cannot be matched to the federal census because they were born after 1930. Individual level data is not publicly available for federal censuses after 1930. As a result, the final linked sample is under-representative 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 blacks in the original sample were born after 1930).

Of those individuals who were born early enough to be found in the federal census, roughly 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

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
<u>Linking child to federal census</u>					
% 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
<u>Linking parents to death certificates</u>					
% 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

is being driven by the white subset of the sample. Nearly 50 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 the 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 slightly 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 initial 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 match the racial distribution of the sample to that of the population. 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. To gain some insight into potential sample selection issues, we have taken a subsample of individuals and transcribed their complete death certificate information regardless of whether they are successfully matched to the census. This allows us to check whether certain individual characteristics are correlated with linking success. Table 2 below presents summary statistics by race and matching success for this subsample.

Several features of the data stand out in Table 2. While age at death is not strongly corre-

Table 2: Summary statistics by linking outcome and race.

Variable	White Individuals		Black Individuals	
	Found in census	Not found in census	Found in census	Not found in census
Age at death	66.3 (16.9)	64.5 (15.3)	56.4 (18.0)	31.8 (27.8)
% male	52.2	57.4	59.5	58.0
% never married	8.2	25.4	16.1	45.0
% born in North Carolina	65.8	57.2	53.2	45.5
% born in South Carolina	13.4	17.7	38.9	47.9
% with agricultural occupation	6.4	4.9	4.8	2.2
% with textile-related occupation	7.3	6.1	1.6	1.8
% with no occupation given	7.7	19.1	10.5	33.2

Notes: Standard deviations given in parentheses. Agricultural occupation includes farmer and farm laborer.

lated 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 concerns about the representativeness of our 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 under-representative 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 our 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 as well as Rosenwaike & Logue (1983) and Hill et al. (2000) on the accuracy of reported ages particularly among older individuals). With noisy measures of longevity, our measures 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. Our data suggest that this is a legitimate concern and that the problem is similar between genders but worse among black individuals relative to white individuals.

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 the 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. 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)).

Table 3: Average age misreporting by gender and race.

	Death certificate birth year - census birth year			Absolute value of (death certificate birth year - census birthyear)		
	Child	Mother	Father	Child	Mother	Father
All	-0.29 (1.58)	-0.58 (2.45)	-0.44 (2.45)	0.98 (1.27)	1.61 (1.93)	1.55 (1.95)
Male	-0.39 (1.47)	-0.57 (2.46)	-0.53 (2.42)	0.96 (1.17)	1.66 (1.89)	1.54 (1.94)
Female	-0.21 (1.62)	-0.59 (2.48)	-0.38 (2.43)	0.98 (1.31)	1.59 (1.99)	1.53 (1.93)
White	-0.47 (1.20)	-0.79 (2.05)	-0.57 (2.20)	0.83 (0.99)	1.41 (1.68)	1.42 (1.77)
Black	0.56 (2.55)	1.08 (4.21)	0.48 (3.85)	1.73 (1.95)	3.39 (2.71)	2.68 (2.94)

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

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 this 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 more reliable than the other, than we would choose to use the more reliable one in our analysis. If they are equally reliable, we could use an average of the two to reduce the variance of the measurement error. 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 or census ages 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. A greater percentage of individuals with birth years or ages ending in zeros or fives would indicate that age is more likely to be being reported with some noise.

Table 4 examines this measure of age heaping for the birth year stated on the death certificate

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	Father's census	Mother's	Child's death	Father's death	Mother's death
	age	age	census age	certificate birth	certificate birth	certificate birth
				year	year	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

and the ages given on the federal census by gender and race.³ There is no evidence of age heaping for children, either on the census or the death certificate. It does appear as if there is age heaping for both father's and mother's ages as reported on the census. Once again, these age reporting issues are more pronounced for the black sample relative to the white sample. The parents' death certificate birth years do not show evidence of age heaping. For this reason, we will use the death certificate birth years in when calculating life spans.⁴

5 Estimates of Intergenerational Correlations in Health

5.1 The Relationship Between Occupation, Longevity and Mobility

Before turning to our measures of mobility based on intergenerational health correlations, it is instructive to compare our health measures to more traditional measures of occupational mobility. By comparing longevity to occupations, we can demonstrate that our measures of longevity are related to socioeconomic status and more clearly establish how our health-based

³It 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. We are assuming 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.

⁴We have used also obtained all of the results using the census birth year and using an average of the census birth year and death certificate birth year. All three approaches lead to similar estimates.

Table 5: Occupational distribution by birth cohort.

<u>Distribution of occupations within cohort</u>				
Decade of birth	Skilled, semi-			
	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.

estimates of mobility relate to the existing literature on intergenerational mobility estimates based on occupations.

Our occupation data come from the occupation question on the death certificates. We 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."⁵ We 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 5 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 the this period while the percentage of skilled and white collar workers both rose significantly. The occupational mobility driven by both

⁵This definition of occupation on the death certificates offers an advantage over the occupation reported on the federal census for studying intergenerational mobility. The census occupation is the individual's current occupation at the time of the census. To the extent that there are occupational transitions over an individual's career, this single observation of occupation is a noisy measure of occupational status over the individual's lifetime. The wording of the death certificate occupation question leads to a measure of occupation that should be a better indicator of an individual's average occupational status over their lifetime and therefore a more useful measure for comparing the occupational outcomes of sons and their fathers.

Table 6: Occupational transitions for sons relative to their fathers.

		<u>Father's occupation</u>				Total
		Farmer	Skilled, semi-skilled	Unskilled	White collar	
<u>Son's Occupation</u>	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

this structural change in the economy as well as more general mobility can be seen in Table 6 giving the occupational transitions of sons relative to their fathers. What emerges from Table 6 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, we can ask whether long term health outcomes differed by occupational category. In particular, we can assess whether the mean life span of an individual differed across occupational category both in terms of the son's occupational category and the father's occupational category. Table 7 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 striking feature of the data that suggests these links is the remarkably short life span for unskilled workers and their children. Sons with an unskilled occupation have a mean life span that is over eight 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

Table 7: Mean life span by occupational category.

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

Notes: Standard deviations are given in parentheses. Lifespan is calculated using the year of death from the death certificate and the year of birth implied by the age given on the federal census.

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.⁶ 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 our 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 seems to be confirmed by the average ages by occupational category.

Table 8 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 son's, the son's occupational categories. A quadratic in child's birth year is included to control for general

⁶A further complication with farmers is that they are the group most likely to live in a rural rather than an urban area. For our time period, there is still a large health penalty to living in an urban area. The long life spans we observe 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.

Table 8: Effects of occupation and occupational transitions on change in longevity, son's life span minus father's life span as dependent variable.

Effects of occupation on life span, child's life span as dependent variable		
	Sons	Daughters
Child's birth year	26.157*** (8.842)	28.860*** (10.879)
(Child's birth year)^2	-0.007*** (0.002)	-0.008*** (0.003)
<u>Son's occupation dummies</u>		
Skilled/semi-skilled	2.818** (1.326)	--
White collar	1.026 (1.243)	--
Unskilled	-8.419*** (2.230)	--
<u>Father's occupation dummies</u>		
Skilled/semi-skilled	-2.260 (1.508)	-2.721 (1.985)
White collar	-2.291 (1.561)	-1.339 (1.793)
Unskilled	-4.494** (2.140)	-4.426* (2.502)
Constant	-24,111.147*** (8,398.513)	-26,560.912** (10,330.568)
Observations	545	411
R-squared	0.44	0.47

Standard errors in parentheses. * significant at 10%; ** significant at 5%; *** significant at 1%. Omitted occupational dummy is farmer for both son and father.

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 contribution of not only an individual's own occupation to longevity but also the occupation of his or her 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

We now turn our attention to estimating correlations in health across generations. Figure 1 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 than 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 throughout 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, we will 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 life span. As Figure 1 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 span's 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. We therefore need to control for cohort in our regressions. We do this by including a quadratic in child's birth year and a quadratic in parent's birth year in the regressions.⁷ In an alternative specification, we include

⁷We allow the coefficients to vary between the polynomial for child's birth year and parent's birth year. This is necessary because the distribution of life spans for children by cohort is different than the distribution of life spans for parents from the same birth cohort. The parental life span distribution is conditional on surviving to child bearing age and having at least one child. This is a very different distribution than the life span distribution for all individuals.

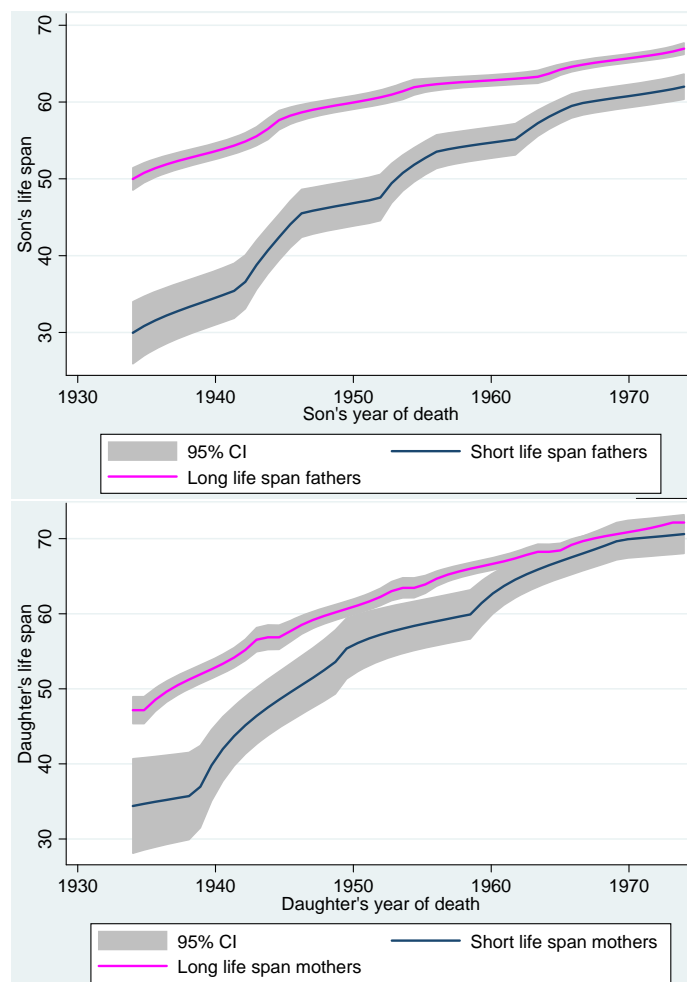


Figure 1: Longevity by cohort, gender and parental longevity. The upper panel is son's longevity across cohorts for father's with life spans shorter their predicted life span based on birth year and for father's with life spans longer than their predicted life span based on birth year. The lower panel is the equivalent graph for daughters based on their mother's life span. Both graphs are kernel-weighted local polynomial fits.

an interaction term between child's birth year and the log of parental life span. This interaction term will capture changes in the intergenerational correlations in longevity over time.

Regression estimates are presented in Table 9.⁸ We run the regressions separately for females and males. To get the largest sample size possible, we focus on only the parent of the same gender as the child. This allows us to use those observations for which we observe only one parent and it makes interpretation of the longevity elasticity simpler by comparing male life spans to male life spans in the son regression and female life spans to female life spans in the daughter regression. Regressions including both parents' life spans are included in Table 10.

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 modern 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, our 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 consistently 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 is biasing these coefficients toward zero. The actual relationship between child and parent longevity is likely stronger than our estimates suggest.

It should be noted that there is also a selection problem that could be biasing our coefficients in the other direction. Among the most recent cohorts in our sample, there is an issue of the sample being under-representative of children who die at young ages but have long-lived parents. Specifically, children dying in 1974 with parents who die after 1974 drop out of the sample because the death certificates are unavailable after 1974. Losing these observations will lead to overestimates of the intergenerational longevity elasticity. In practice, this selection problem appears to be minor. First, it only occurs when parents outlive their children, meaning that the parents have a life span that is greater than their child's plus the parent's age when the child was born. This typically means the parent would have a life span 20 to 30 years longer than the child's, a rare occurrence in the earlier cohorts for which we could observe such a gap.

⁸Regressions using life span rather than the log of life span are presented in the appendix.

Table 9: Intergenerational longevity elasticity regressions, log of child's life span as dependent variable.

	Sons			Daughters		
	(1)	(2)	(3)	(4)	(5)	(6)
ln(parent's life span)	0.283*** (0.063)	0.274*** (0.066)	0.214*** (0.075)	0.185*** (0.061)	0.203*** (0.064)	0.210*** (0.068)
Child's birth year	0.995*** (0.183)	0.692*** (0.245)	0.634** (0.247)	0.795*** (0.176)	0.760*** (0.274)	0.772*** (0.277)
(Child's birth year)^2	-0.000*** (0.000)	-0.000*** (0.000)	-0.000*** (0.000)	-0.000*** (0.000)	-0.000*** (0.000)	-0.000*** (0.000)
Parent's birth year		0.362* (0.190)	0.257 (0.200)		0.053 (0.246)	0.080 (0.258)
(Parent's birth year)^2		-0.000* (0.000)	-0.000 (0.000)		-0.000 (0.000)	-0.000 (0.000)
ln(parent's life span) x (child's birth year-1900)/10			0.093* (0.056)			-0.019 (0.055)
Constant	-930.676*** (174.020)	-980.706*** (175.897)	-789.322*** (210.251)	-737.596*** (167.016)	-754.574*** (170.256)	-799.105*** (215.053)
Observations	586	586	586	425	425	425
R-squared	0.38	0.38	0.38	0.41	0.41	0.41

Notes: Standard errors in parentheses. * significant at 10%, ** significant at 5%, *** significant at 1%. All life spans are calculated by subtracting year of birth on the death certificate from year of death given on the death certificate.

The father's life span is used for parent's life span in the case of sons and the mother's life span is used in the case of daughters.

Second, examining the probability of matching parents to death certificates by year of child death reveals very similar match rates for children dying in 1974 as for children dying in earlier decades suggesting that this selection issue is not having a significant impact on the sample and lowering match rates for the 1974 cohort. While this selection problem should not be dismissed, the impact on the estimated coefficients are likely small in comparison to attenuation bias from age misreporting.

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 .12 compared to an elasticity of .31 for a son born in 1910. For daughters, the coefficient on the interaction term is actually negative but quite small 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 10 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 son's, 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 is 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.

6 Extensions

These initial results on the intergenerational correlations in life span are encouraging evidence that health measures may be a practical way to study historical mobility rates in a way that can

Table 10: Intergenerational longevity elasticity regressions including both parents.

Dependent variable:	ln(child's life span)		child's life span	
	Sons	Daughters	Sons	Daughters
ln(father's life span)	0.359*** (0.082)	0.090 (0.095)	-- --	-- --
ln(mother's life span)	0.157** (0.076)	0.320*** (0.086)	-- --	-- --
Father's life span	-- --	-- --	0.202*** (0.058)	0.055 (0.072)
Mother's life span	-- --	-- --	0.056 (0.054)	0.188*** (0.068)
Child's birth year	1.516*** (0.279)	0.805** (0.362)	53.497*** (13.164)	32.009* (18.316)
(Child's birth year)^2	-0.000*** (0.000)	-0.000** (0.000)	-0.014*** (0.003)	-0.009* (0.005)
Constant	-1,427.452*** (265.344)	-749.339** (343.910)	-50,149.391*** (12,507.161)	-29,573.039* (17,398.078)
Observations	293	215	293	215
R-squared	0.41	0.41	0.43	0.45

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

make meaningful distinctions between the experiences of males and females. The preliminary results presented in this paper suggest several possible extensions both in terms of fairly simple extensions that can be executed with the current data and minor additions to the dataset as well as more ambitious projects that would require more substantial investment but seem promising given the outcomes of this pilot project.

Currently the dataset is being extended both in terms of the geographical coverage of the sample and the density of the sample. This 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 would allow us to investigate 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, a larger sample will allow us to begin to explore racial differences in the intergenerational transmission of health status.

A second extension currently under way is to better utilize 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 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 2 depicts these adjusted healthy life spans relative to actual life span for our current sample. This measure of healthy life span would offer a measure of long run health that may be a better measure of overall welfare than the actual life spans that are focus of this paper. Coming up with a way to categorize the various conditions and their impact on quality of life could further refine our 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 our sample to their adult households in the federal census, an additional set of interesting variables could 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. Spouse's occupation would make it possible to construct occupational mobility measures for daughters comparable to those used

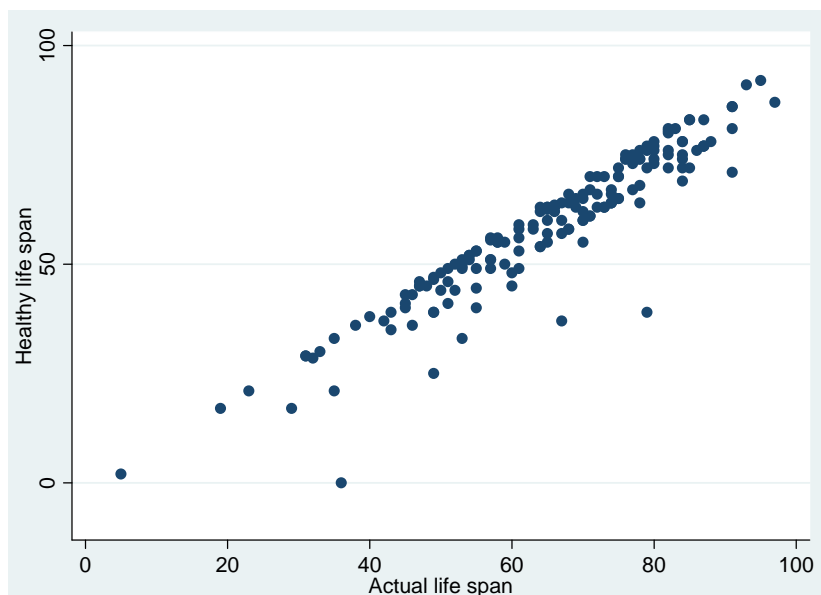


Figure 2: Healthy life span versus actual life span. Healthy life span is calculated as actual life span minus years spent in poor health where years in poor health come from the durations of significant medical conditions reported on the death certificate.

in Ferrie (2005) and Long & Ferrie (2007) for sons 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, age at first pregnancy, the number and spacing of children, labor force participation decisions and spouse characteristics.

7 Conclusion

This paper has introduced a unique intergenerational dataset of death certificate data that allows for estimating intergenerational correlations in long run health outcomes. These measures of health across generations 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 unique features of name reporting on death certificates also allows for extending the census linking strategies of other mobility studies to females.

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. We 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 sons in skilled, semi-skilled, white collar or agricultural jobs. 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. These correlations are stronger between parents and children of the same gender. The elasticities for daughters have appears relatively stable over time while there is evidence of a gradual increase in the strength of the correlation between father and son's life spans over the past century consistent with the previously established decline in occupational mobility.

These results suggest that long run health outcomes are 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.
- Abbot, M., Abbey, H., Boiling, D., & Murphy, E. (1978). The familial component in longevity-a study of offspring of nonagenarians: 111 Intrafamilial studies. *American Journal of Medical Genetics*, 2, 105–120.
- 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).
- 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.
- Mazumder, B. (2005). Fortunate sons: New estimates of intergenerational mobility in the United States using social security earnings data. *Review of Economics and Statistics*, 87(2), 235–255.
- 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.
- Pearl, R., & Pearl, R. (1934). *The ancestry of the long-lived*. Johns Hopkins press.

- Rosenwaik, 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.
- 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). *Eighth 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.

A Additional Tables

B Death Certificate Variables

The format of the North Carolina death certificates has changed over time. Below 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

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

	Sons			Daughters		
	(1)	(2)	(3)	(4)	(5)	(6)
Parent's life span	0.177*** (0.044)	0.181*** (0.046)	0.179*** (0.050)	0.129*** (0.046)	0.144*** (0.048)	0.153*** (0.049)
Child's birth year	33.319*** (8.335)	24.618** (11.155)	24.548** (11.178)	31.868*** (8.616)	29.103** (13.404)	29.715** (13.426)
(Child's birth year) ²	-0.009*** (0.002)	-0.007** (0.003)	-0.007** (0.003)	-0.009*** (0.002)	-0.008** (0.004)	-0.008** (0.004)
Parent's birth year		10.950 (8.607)	10.558 (9.133)		3.961 (11.999)	7.643 (12.718)
(Parent's birth year) ²		-0.003 (0.002)	-0.003 (0.002)		-0.001 (0.003)	-0.002 (0.003)
Parent's life span x (child's birth year-1900)/10			0.005 (0.038)			-0.036 (0.041)
Constant	-30,948.801*** (7,916.727)	-32,914.887*** (8,012.306)	-32,447.435*** (8,799.051)	-29,433.062*** (8,173.213)	-30,518.561*** (8,327.704)	-34,796.888*** (9,658.436)
Observations	586	586	586	425	425	425
R-squared	0.43	0.44	0.38	0.48	0.48	0.48

Notes: Standard errors in parentheses. * significant at 10%, ** significant at 5%, *** significant at 1%. All life spans are calculated by subtracting year of birth on the death certificate from year of death given on the death certificate.

The father's life span is used for parent's life span in the case of sons and the mother's life span is used in the case of daughters.

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