"The Past as Prologue: The Effect of Early Life Circumstances at the Community and Household Levels on Mid-Life and Late-Life Outcomes"

Joseph Ferrie, NORTHWESTERN UNIVERSITY
Karen Rolf, UNIVERSITY OF NEBRASKA
Werner Troesken, UNIVERSITY OF PITTSBURGH

March 7, 2008

Abstract

We explore the impact of circumstances early in life on later life outcomes in the U.S. using data linking individuals from the manuscript schedules of federal population censuses shortly after their birth to Social Security records and state death records. This allows us to assess the effect of individual, household, and community level influences on education, income, longevity, and cause of death. In our preliminary analysis, we find that season of birth, parents' literacy and fluency in English, and the local mortality environment during the first five years of life had substantial affects on age at death, as did characteristics measured at enlistment into the U.S. Army in World War Two. The current paper extends this work by examining measures of education and income (mid-life outcomes), as well as cause-specific mortality, and how they relate to early life circumstances.

Introduction

Health outcomes are among the most intractable forms of inequality in modern society. Though inequalities in income or education or access to housing can be addressed with various forms of redistribution in the very short run, health is the product of a lifetime of influences. Modern research emphasizes the extent to which events very early in life – even events *in utero* – can have a significant impact on health much later in life. We follow a

large population of more than 40,000 individuals born between 1900 and 1930 from their appearance in the U.S. federal population censuses until their deaths and their appearance in the Social Security Death Index (SSDI) and in state death records to assess the impact of early life-circumstances on longevity and cause-specific mortality, as well as how the impact of early life circumstances has evolved across these 30 years of birth cohorts.

Though other studies have examined early-life influences on later-life outcomes, until now none have done so with such detailed information on early-life circumstances at the individual, family, and neighborhood levels, or done so with a study that is prospective in design. Our work does not suffer from the individual's inexact recollection of their own early-life circumstances, nor does it suffer from the attrition bias that occurs in long-run longitudinal studies. The large number of birth cohorts we examine make it possible to assess how early-life influences changed from the beginning of the twentieth century until the early Great Depression. This allows us to more accurately project the longevity and late-life health of cohorts that are now entering their mid-seventies, and thereby evaluate the extent to which health inequality will remain a concern for policymakers over the coming decades.

The impact of circumstances early in life on health outcomes late in life has been a subject of increasing interest since the 1970s.[1] Particularly since the work of Barker and a number of co-authors, attention has focused on the environment faced by individuals not just in the years immediately after birth but even in utero in shaping their health decades later. A shortcoming of much of the research in the area is the limited range of information

[1] See Gabriele Doblhammer "The Late Life Legacy of Very Early Life." Max Planck Institute for Demographic Research Working Paper WP2003-030, September, 2003, for a survey of the literature.

available on the early life experiences of individuals whose later life health can be observed: no modern longitudinal datasets span more than 40 years. Information is often collected retrospectively, or small, opportunistic samples that may not be very representative are exploited.

We offer two improvements on previous work: (1) a large sample (2) that is created prospectively. The sample allows us to assess the impact on one particular late life health outcome, longevity, of circumstances both very early in life and at approximately age 25. Recent research by Karen Clay and Werner Troesken has established a relationship between the environment (crowding and impure water) and health fifteen years later at the city level for the early nineteenth century U.S.[2] We will examine the same time period, as well as outcomes that occur as late as 2005, and focus on circumstances that can be observed not just at the community level but also at the household and individual level.

The Data

The data was created in two steps: (1) first, 1,537,659 death certificates from eight states were obtained (see Figure 1), from which we randomly drew 96,099 males who were under age 5 in 1920; (2) then, these individuals were sought in the manuscript schedules of the 1920 U.S. Census of Population and in the World War Two enlistment records of the U.S. Army. A total of 28,839 males (30%) were linked from the death records to the census records; of these, 5,836 were also linked to the army records. The linkage rate of 30% from the death records to the census results from individuals missed or incorrectly enumerated in

[2] "Deprivation and Disease in Early Twentieth Century America." NBER Working Paper No. 12111, March, 2006.

the census or individuals who could be matched to more than one person in the 1920 Census and will be improved upon once additional information from Social Security records (exact place of birth, full names of both parents) is added.

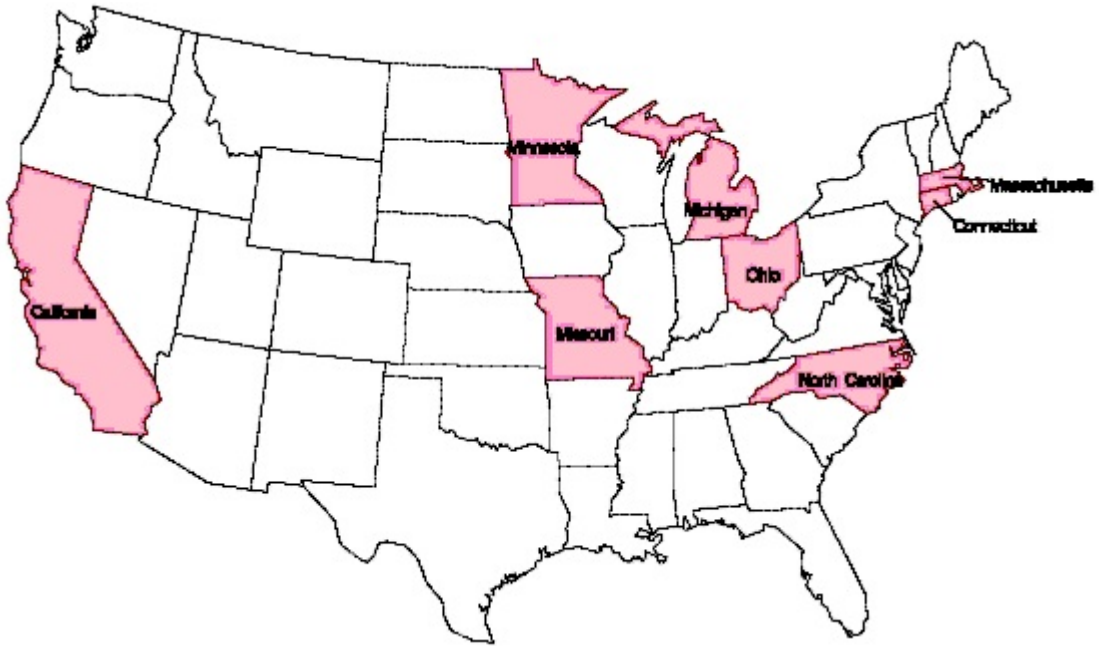

Figure 1. State From Which Death Certificates Used Were Obtained.

The states used for this preliminary analysis were selected because of the easy availability of their computerized death records. But they are also useful because all were members of the Death Registration Area by 1920, so we will be able to attach city-level mortality statistics for the years prior to 1920 to each record. These states also include nine of the country's twenty largest cities in 1920.

The census data, together with information from the state death certificates on cause-specific mortality by state and year, provide information on circumstances before and after birth; the enlistment data provide information on subsequent health (measured by Body

Mass Index or BMI); finally, the state death records and Social Security records make it possible to calculate the age at death for members of the sample. We also have measures of education (from the state death records) and income (inferred from the monthly Social Security benefit).

Shortcomings of the Data

The present analysis uses mostly males because the linkage process uses name and date and place of birth, but women's name changes at marriage prevent their linkage. But Social Security is providing information on women's names at birth to help, and state death records post-1978 provide maiden name. Cause of death info (along w/education) will come from 50 sets of state death records (8 are now in hand: CA, CT, MA, MI, MO, MN, NC, OH).

The Social Security Death Index is available only 1965-2005, so the "window" in which we can observe deaths is only 40 years of calendar time (though 70% of the 1910-1919 birth cohort died in this window). As a result, for each cohort, we will need to limit the ranges for age at death within which we examine the correlates of mortality. For example, for the sample drawn from the 1920 census (males born 1915-1919), we can look at only those who died between 50 and 85. To examine the correlates of longevity, we will run regressions of the form:

$$E(Age^{death} \mid Age^{death}_{min} < Age^{death} < Age^{death}_{max}) = \beta' X_i + \gamma' Y_i + \delta' Z_i + \varepsilon_i$$

where X_i are individual & household characteristics, Y_i are neighborhood characteristics and Z_i are economy-wide effects (e.g. GDP, pandemic, war).

State death records are only generally available 1970-2006, so the "window" there is even smaller. California's records, however, go back to 1940, and also make it possible to include women (most states do not report birth surnames until 1979): California birth records are also computerized from 1905, so by matching on birth date, given name, and birthplace (CA) it was possible to link birth and death records to census records for 25,000 males and females who died in California. This will make it possible to assess how much information is lost by the smaller window within which deaths are observed in other states and by the focus on males in other states.

The World War Two data has some oddities. Individuals are selected on the basis of physical fitness for military service, so their mortality after the war is somewhat better than the general population over the first two the decades after war. Their height and weight reflect the selection criteria in place at enlistment (which change over war). The military provided tobacco, leading to higher than average lung cancer and heart disease at older ages Finally, the data on height and weight are only available for enlistments between July, 1940 and February, 1943.

Data Analysis

The outcomes we can currently analyze are longevity and height and weight at enlistment of individuals born 1915-1919 and linked from State Death Records to the 1920 U.S. Census and WWII records. Why have we focused here on 1915-1919 births and 1920 Census? Because this maximizes the number of links to WWII records and to State Death Records (70% of this cohort dies 1970-2006) . It also allows us to shows effect of conditions

for individuals under age 5, and the impact of 1918-19 influenza pandemic.

In Table 1, the outcomes examined are age at death as a function of early life household circumstances and physical characteristics (height, weight, BMI) around age 25 (conditional on enlisted in World War II and dying after 1970 in one of the eight states from which death records were drawn).Only males are included because, at present, the linkage process requires that the individual's surname did not change between their appearance in the pre-1940 census and their appearance in the Social Security records or state death records. Later work will overcome this shortcoming by exploiting information on the names of each decedent's parents in the larger set of Social Security records from which the Death Index is abstracted and in the state death certificates. We will also eventually add the information on exact cause of death, as well as the amount of the individual's Social Security monthly benefit – a good proxy for permanent income in the years before retirement.

Column 1 examines the influences on longevity, measured in days, conditinal on death between ages 67 and 83. The most noteworthy of the early household characteristics is race (whites lived more than 8 months longer) and the absence of a father in the 1920 household (such individuals lived 6 months less than those whose fathers were present). The regressions also included controls for state of residence in 1920, and year and month of birth. Figure 2 shows the pattern of month-of-birth effects: the starkest difference is between those born in August (who live to 72.6 years) and those born just 2 months earlier in June (who live to 71.4 years). We do not at this point have a simple explanation for this result, though it is consistent with findings by Doblhammer (see footnote 1), who suggests it is related to conditions *in utero* at crucial gestational ages.

Variable	Age at Death (days)	Height at Enlistment (inches)	Weight at Enlistment (pounds)	BMI at Enlistment
Household Size	-11.485	-0.061**	0.175	0.070**
	15.817	0.031	0.258	0.034
Father Absent	-172.520**	-0.131	-4.675***	-0.588***
	80.221	0.180	1.481	0.197
Mother Absent	14.764	-0.600***	0.226	0.452**
	93.255	0.206	1.692	0.225
Birth Order	1.044	0.006	-0.391	-0.067*
	17.308	0.035	0.286	0.038
Died in State of Birth	64.719*	-0.101	1.069*	0.234***
	36.909	0.078	0.643	0.086
White	265.228***	0.274*	-0.054	-0.156
	71.885	0.161	1.319	0.175
Top 50 Cities, 1920	-18.655	-0.071	1.606**	0.279***
	38.987	0.081	0.666	0.089
Controls				
Year of Birth	Yes	Yes	Yes	Yes
Month of Birth	Yes	Yes	Yes	Yes
State, 1920	Yes	Yes	Yes	Yes
Observations	24,993	5,836	5,836	5,836
Adjusted R^2	0.024	0.030	0.009	0.014

Standard errors in parentheses. * p<0.10 ** p<0.05 ***p<0.01

Table 1. OLS Regressions on Age at Death (ages 57-83), and Height, Weight, and Body Mass Index (BMI) at Enlistment in World War II. U.S.-born males only.

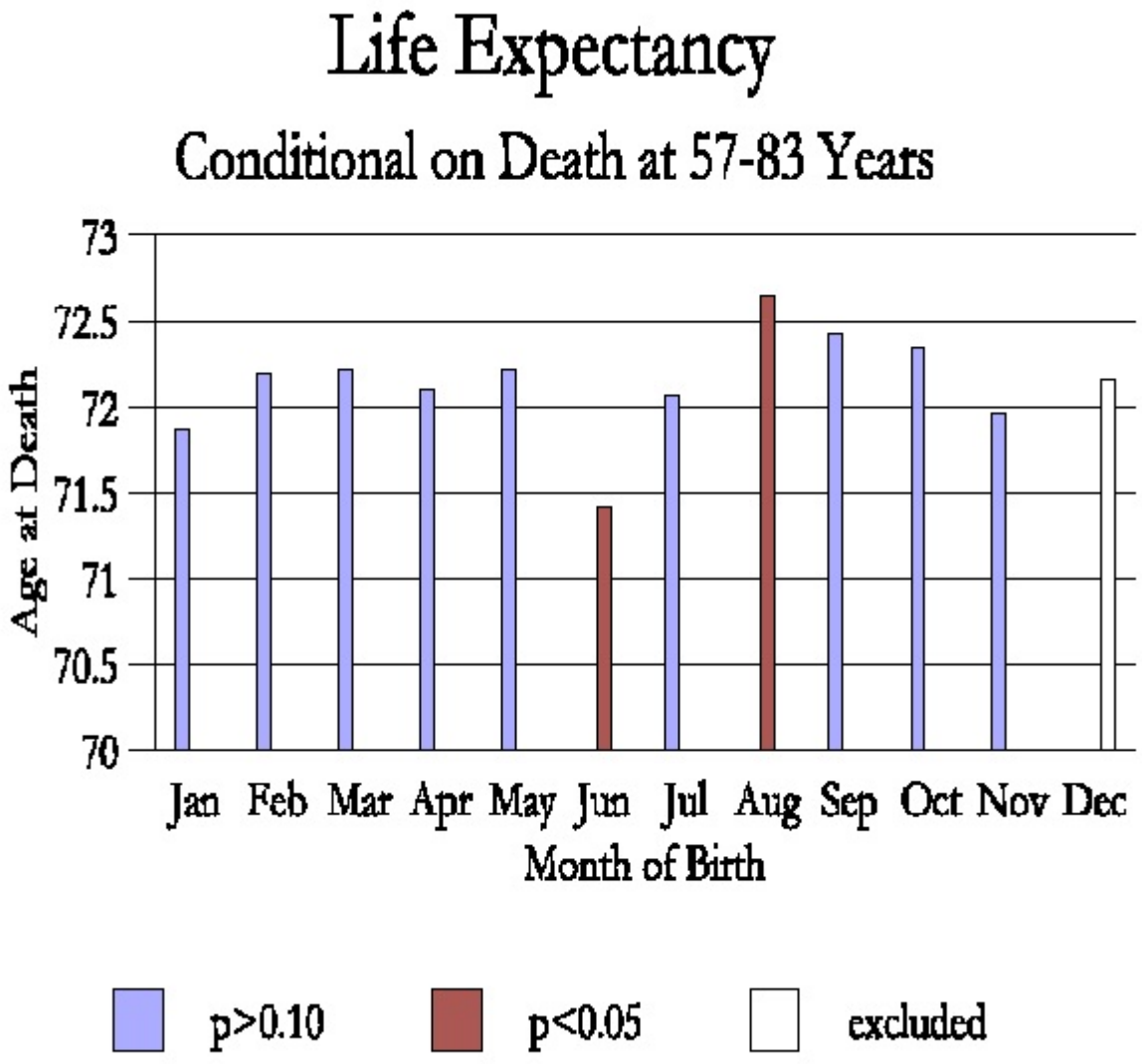
Life Expectancy
Conditional on Death at 57-83 Years
Age at Death
73
72.5
72
71.5
71
70.5
70
Jan Feb Mar Apr May Jun Jul Aug Sep Oct Nov Dec
Month of Birth
p>0.10
p<0.05
excluded

Figure 2.

The next two columns in Table 1 examine the impact of early-life circumstances on the individual's physical appearance at enlistment in world War II. Individuals from larger households and families in which the mother was absent were shorter on average than otherwise identical individuals. The "absent mother" effect is particularly large, amounting to more than a half inch. As stature is thought to be a reasonable proxy for net nutritional status (calories ingested minus calories consumed for basal metabolism, fighting disease, and physical activity) over the early years of life, this effect might be picking up either poor nutrition or poor disease immunity early in life if the mother's absence began soon after birth. We will eventually compare this effect for individuals whose mothers died or left the family soon after their birth and for individuals who were several years older when the absence began, in order to isolate the mechanism through which it operates.

Weight at enlistment was also related to early life circumstances, though the specific characteristics of the individual's home environment that mattered were different from those that influenced height. For weight, an absent father rather than an absent mother was detrimental. Since weight is a better measure of contemporary nutritional and health status than height, this may indicate that individuals whose father was absent early in life have entered the labor force at earlier ages than those whose fathers were present, leading to higher caloric expenditures and perhaps greater exposure to new disease environments. Weight was also higher among those resident in the 50 largest cities in 1920, which may indicate the range of food products available at reasonable prices, particularly compared to smaller towns and rural places dependent on staple crops.

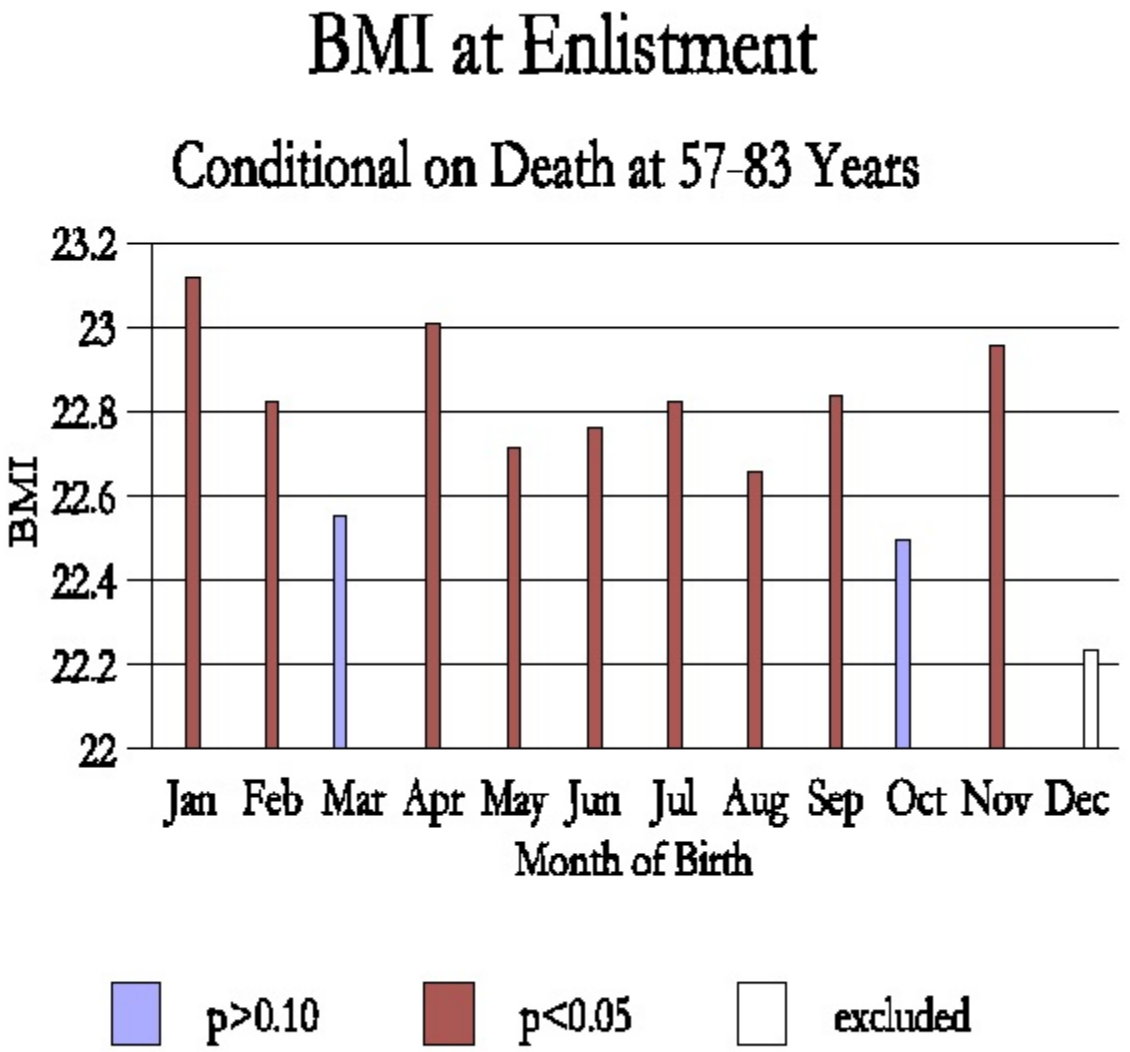

Figure 3.

As was the case with longevity, there are season-of-birth effects on BMI at enlistment: there are local peaks in January, April, and November, and local troughs in March, October, and December. The effect of year of birth is also pronounced: individuals who were born just before and during the influenza pandemic had lower BMI at around age 25 than those born in 1915.

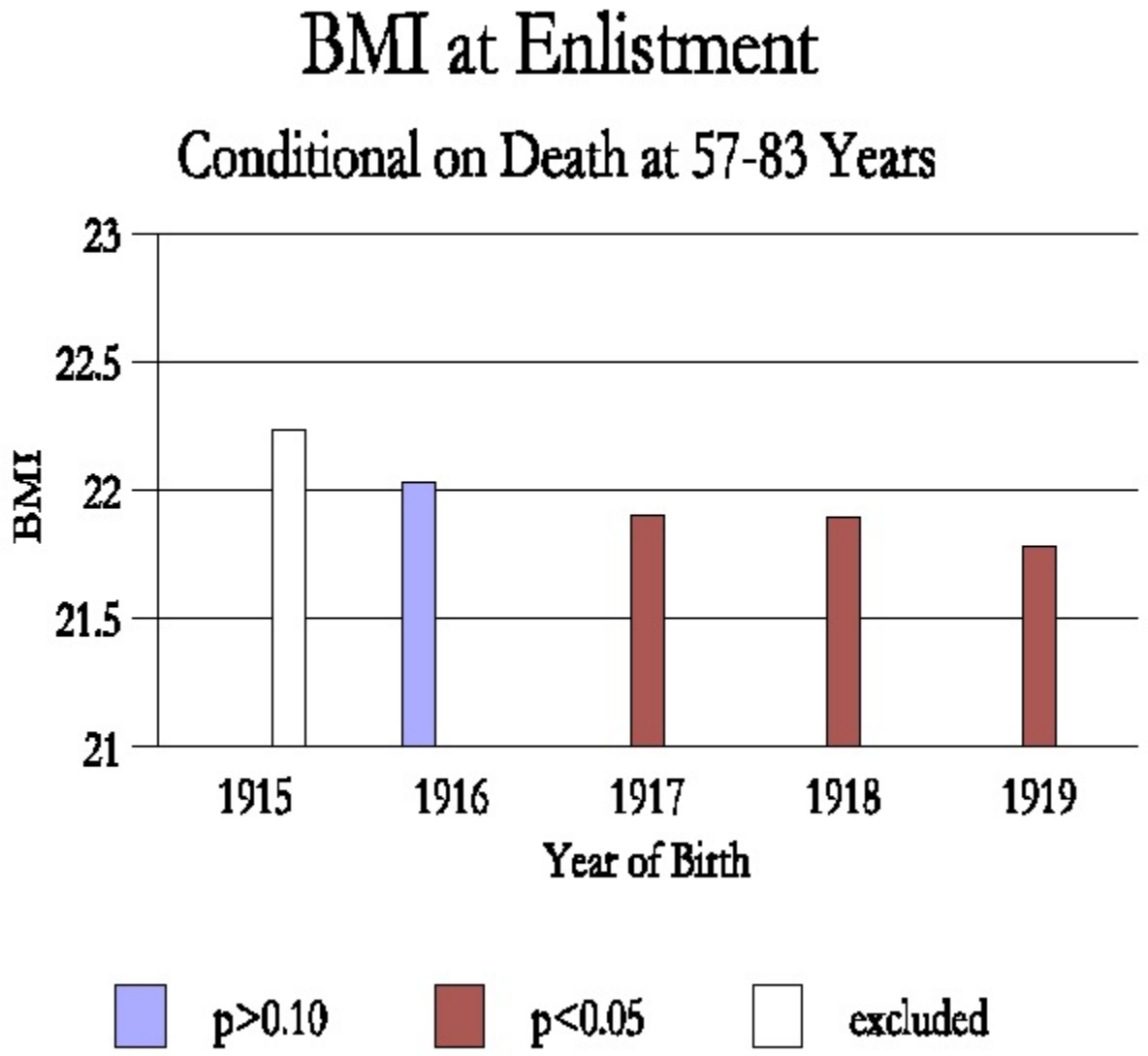

Figure 4.

Finally, we can use the information on both early life circumstances and circumstances at age 25 to predict longevity, by adding height, weight, or BMI to the regression in column 1 of Table 1. Figure 5 shows the result for the addition of height and height squared: longevity is maximized by a height of 70.5 inches at age 25.

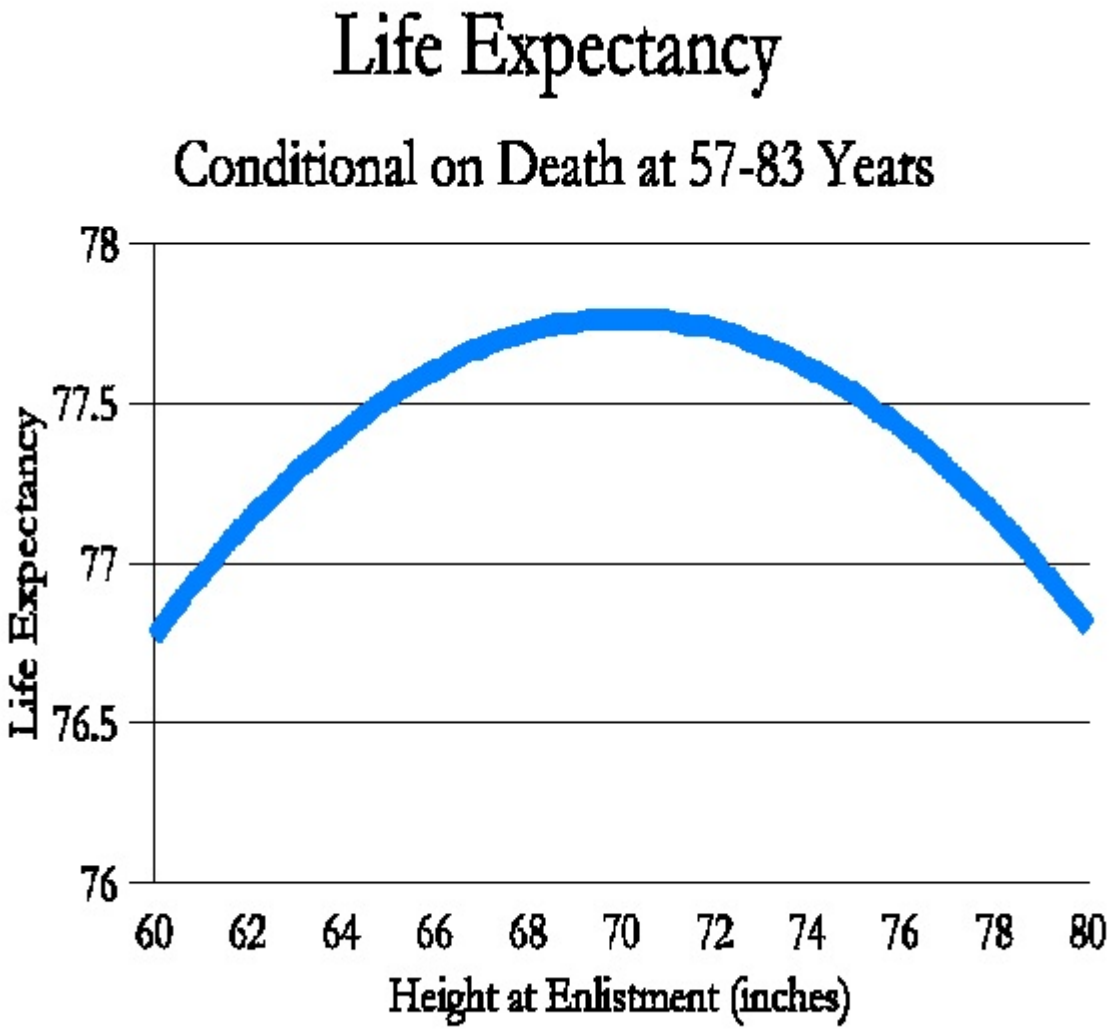
Life Expectancy
Conditional on Death at 57-83 Years
78
77.5
77
76.5
76
Life Expectancy
60
62
64
66
68
70
72
74
76
78
80
Height at Enlistment (inches)

Figure 5.

Sensitivity Analysis

The California data (births in CA 1915-1919 linked to deaths in CA 1940-2004, using only given name, exact date of birth, and place of birth/death – CA) makes possible two additional comparisons:

1. males vs. females: does the impact of early life conditions differ by *sex*?

and

2. how much of the impact for *males* is missed because of the 1970-2006 "window" for the State Death Records outside CA?

Table 2 shows regression results for males and females age 26-85 at death, and for males age 25-60 and 61-85 at death. The most striking difference from the results in Table 1 (where the age at death was limited to 57-83) when the age range is expanded to 26-85 is the change in the season of birth effect: June is now the best month to be born, and October is the worst (a June birth adds roughly 2 years to life span compared to an October birth). The absent father effect is again present, though nearly double its effect in Table 1.

When males and females are compared (column 1 vs. column 2), the general conclusion that emerges is that males are more susceptible to season of birth, but females show a larger (negative) impact from residence in large cities and being born during the influenza pandemic (females born in 1918 lived roughly a month longer than females born just before the outbreak). When younger (age 25-60) and older (age 61-85) deaths are compared (column 3 vs. column 4), the effect of season of birth is limited to those who died at older ages, while the effect of being born around the time of the influenza pandemic

Variable	Males 26-85 (days)	Females 25-85 (days)	Males 25-60 (days)	Males 61-85 (days)
Constant	24538.000***	25148.000***	17201.000***	27157.000***
Birth Month				
Jan	1.663	-208.183	-154.303	21.335
Feb	371.179*	391.676	401.440	-2.076
Mar	100.429	-220.752	-5.188	-29.669
Apr	253.270	258.225	151.499	98.838
May	-60.844	-95.133	58.700	69.378
Jun	323.505*	-314.428	315.869	-6.087
Jul	171.370	29.344	290.171	7.820
Aug	216.362	-7.954	119.487	94.809
Sep	210.095	-157.530	414.505	117.448
Oct	-331.780*	-247.663	-7.337	-190.580*
Nov	-5.289	246.463	-257.893	-49.973
Dec	–	–	–	–
Birth Year				
1915	–	–	–	–
1916	-171.901	350.269**	-383.184**	23.131
1917	7.283	471.441***	-422.503***	94.780
1918	120.254	438.936***	-289.072	54.760
1919	-75.760	481.811***	-519.984***	-49.056
Household Size	41.786	36.251	-32.386	24.070
Father Absent	-316.077*	-81.759	-516.014**	4.499
Mother Absent	-88.656	342.034	122.863	-112.612
Birth Order	-30.095	-26.473	147.556***	-37.393
Top 50 Cities, 1920	-19.931	-331.200***	376.439***	-55.010
Observations	15,328	7,809	3,599	11,729
Adjusted R^2	0.001	0.002	0.007	0.000

* p<0.10 ** p<0.05 ***p<0.01

Table 2. OLS Regressions on Age at Death (25-85, 25-60, & 61-85), Males and Females Born and Died in California, 1940-2004.

shows up only for those who died between 25 and 60. Identifying the mechanisms through which these effects operated will be easier when we are able to examine exact causes of death.

Future Directions

Additional information on each individual's 1920 household will be gathered from the manuscript census pages: father's occupation, parents' literacy, the exact location (street address) of the household, and the characteristics of adjacent households. With this information, it will be possible to assess the impact of socioeconomic status on longevity. The information on street address will be particularly useful. Figure 6 shows Toledo, Ohio, and the individuals in the sample who were located there in 1920. Red stars are those who died earlier than predicted by the regression in Table 1; green stars are those who died later than predicted. Detailed local area maps will allow us to examine whether proximity to environmental hazards (factories, mills, gas stations, railroads) helps account for these residual effects. Data on local mortality rates from published totals by city in the Death Registration Area will allow us to examine the impact of the disease environment in which the individual grew up on later outcomes.

We will also add more detail on later life outcomes, in addition to specific cause of death: Social Security retirement benefits, applications for Social Security disability payments, occupation at death, and educational attainment will be available for more than half of those now linked.

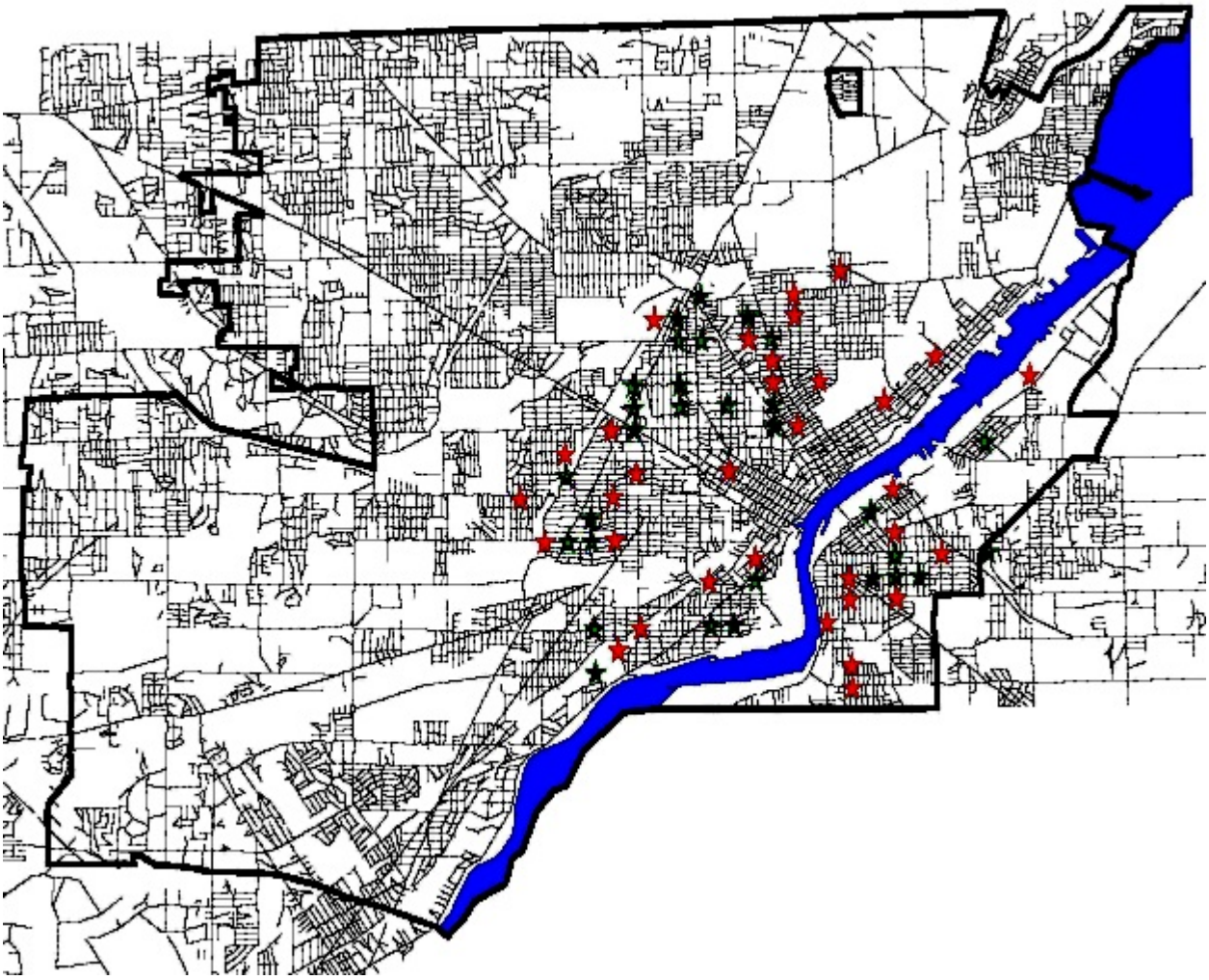

Figure 6.

Conclusions

We have been able to examine the impact of early life circumstances on later life health for the first time with a large, representative population free from attenuation or recall bias. The most important results are:

1. Month of birth matters, but the effect varies by age and sex;

2. Home environment (presence of parents, household size, birth order) matters

3. The effects of early environment are stronger for males than for females (greater male *frailty*?);

4. Absent father → shorter life, lower weight & BMI, but Absent mother → shorter stature;

5. The strongest impact of 1918-19 influenza pandemic is on weight & BMI at enlistment in WWII;

6. Height around age 25 has a non-linear effect on longevity (optimum=70.5 in.);

7. At least in California, early life conditions (except for month of birth) have a *stronger* effect on longevity at *younger* ages → the effects we find using data from other states at *older* ages is probably a *lower bound*.