"… *Healthy, Wealthy, and Wise?: Later Life Physical, Economic, & Cognitive Effects of Early Life Circumstances in the U.S.*"

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

September 11, 2008

Extended Abstract

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. Particularly since the work of Barker and a number of co-authors (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), 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 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 ("Deprivation and Disease in Early Twentieth Century America." NBER Working Paper No. 12111, March, 2006) 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. 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 dataset was created in two steps: (1) first, 137,224 males who were under age 5 in 1900, 1920, and 1930 and under age 2 in 1910 were drawn from the 1900-30 IPUMS files (Steven Ruggles *et al.*, Integrated Public Use Microdata Series: Version 3.0 [Machine-readable database]. Minneapolis, MN: Minnesota Population Center [producer and distributor], 2004) – large, nationally-representative samples extracted from the U.S. decennial population censuses; (2) then, these individuals were sought in the Social Security Death Index (which reports the name, date of birth, and date of death for individuals who were in the Social Security system between 1965 and 2005) and in state death certificates and in the World War Two enlistment records of the U.S. Army.[1] A total of 18,195 males were linked from one of the pre-1940 censuses to the Social Security records; of these, 1,218 were also linked to the army records. The linkage rate of 13.3% from the IPUMS to the SSDI represents roughly a quarter of those who – based on survivorship by age of the white, native-born male U.S. population – were still alive in 1965 and therefore at risk to enter the Social Security Death Index.[2]

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 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).

Preliminary Analysis

In Table 1, we present an example of our preliminary analysis of the sample. The outcome examined is age at death as a function of both early life circumstances (conditional on having survived to enter the Social Security system and having died by November, 2005 when the Social Security file ends).[3]

---

[1] 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. 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.

[2] The greatest source of linkage failure, apart from the absence of anyone at all with the correct name, birth month, and birth year in the SSDI, was the inability to differentiate among several individuals who were identical along these three dimensions. At a later stage in the project, additional information will be provided by the Social Security Administration (the full names of both parents, and the detailed place of birth) that will make it possible to resolve these ambiguities. The linked data now also contains an unknown number of "false positives" (individuals from the IPUMS matched to the wrong individual in the SSDI), but these, too, will be eliminated when the additional data are added. The only substantial bias introduced by the inability to resolve ambiguous links is that the sample is biased toward those with more unusual combinations of name, birth year, and birth month.

[3] Table 1 presents Ordinary Least Squares (OLS) regressions. All of the substantive findings that follow were also generated when a hazard model was estimated instead.

Among individuals who survived to age 70 and died by age 95 (Columns 1 and 2), longevity was increasing in year of birth, but was sharply lower among those born in the second quarter of the year and somewhat lower among those born in the third quarter. This is broadly consistent with work by Doblhammer (2003) who also found that births in the second quarter of the year resulted in shorted lives than births in the fourth quarter of the year, though we are able to control for a much larger set of early life circumstances. The life spans of those born in April, May, and June were nearly 31 weeks shorter than those in the excluded category (born October to December). We are investigating the source of this difference. The household mortality rate (one minus the ratio of children surviving to children ever born to the mother) is also associated with shorter lifespans (Column 2). This could indicate circumstances within the household (either genetic or environmental), or it could reflect the broader community-level health environment. When both the state-level and household-level measures are entered, the latter has a large, negative impact on longevity. The effect of quarter of birth, however, persists. Among individuals who died between ages 50 and 85 and were observed in the 1920 censuses, the only significant impact on longevity comes from residence in the South (a reduction of nearly half a year compared to those in the Northeast) and family size.

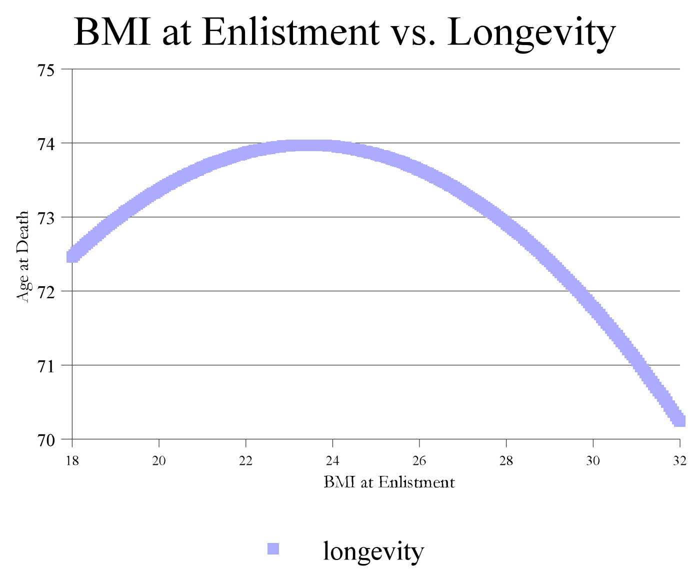

Figure 1.

The elimination of the quarter of birth effect in going from Column 1 and 2 to Column 3 could result from differences across cohorts in the impact of early life circumstances on mortality, or from differences in the range of age at death: if early circumstances have an impact only on causes of death for which individuals are at heightened risk only after having survived causes of death not so sensitive to early circumstances, the pattern we observed would be expected. Though Column 3 reveals no substantial penalty paid in longevity terms by those who were born during the influenza pandemic of 1918-19, additional analyses (not shown) which include BMI at enlistment in World War Two as a regressor find an individual born in 1918 survived more than two years fewer than one born in 1915. This is consistent with the long-lingering effects of exposure to the pandemic described by Douglas Almond ("Is the 1918 Influenza Pandemic Over? Long-Term Effects of *In Utero* Influenza Exposure in the Post-1940 U.S. Population," *Journal of Political Economy* 114 (2006): 672-712). The effect of BMI on longevity is non-linear: for example, in Column 5, longevity achieves a maximum at a BMI of 24. Figure 1 plots this relationship.

The analysis we are preparing for the 2009 IUSSP meetings expands on this work and incorporates three substantial improvements: (1) the sample will be expanded to more than 400,000 individuals, including females by making use of the new 5% Public Use Samples from the 1900 and 1930 U.S. population censuses, which will make possible both an examination of the impact of the characteristics of neighboring households and proximity to environmental hazards early in life on longevity and an examination of how males and females differ in how early life circumstances influence their subsequent health; (2) mid-life characteristics (IQ, educational attainment and income imputed based on the monthly Social Security benefit received by each individual) will be included as both outcomes generated by early life circumstances and as predictors of longevity; and (3) the link between circumstances early in life and the specific cause of death will be examined.

Table 1. OLS Regression on Age at Death with Mortality Measures

	(1) 1900-10, Age 70-95	(2) 1900-10, Age 70-95	(3) 1920, Age 50-85
Midwest	-0.090	-0.933	-0.681*
	(0.44)	(1.61)	(1.66)
South	-0.329	-2.091*	-1.324***
	(1.48)	(1.86)	(3.06)
West	-0.180	-1.345	0.143
	(0.50)	(1.50)	(0.26)
City Pop. 100,000+	-0.313	-0.519	-0.081
	(1.40)	(1.14)	(0.25)
City Pop. 10,000-99,999	-0.154	0.537	-0.087
	(0.61)	(0.99)	(0.24)
Year of Birth	0.076***	0.092	
	(5.14)	(1.14)	
Born Jan-Mar	-0.258	0.984*	0.218
	(1.19)	(1.83)	(0.56)
Born Apr-Jun	-0.623***	-1.540***	0.458
	(2.81)	(2.72)	(1.21)
Born Jul-Sep	-0.398*	0.213	0.195
	(1.85)	(0.37)	(0.51)
Family Size	0.036	0.003	-0.122*
	(0.94)	(0.03)	(1.82)
Household Mortality Rate	-0.713	-5.019***	
	(1.06)	(2.77)	
State Mort Rate at Birth		-0.160	-0.143
		(1.20)	(1.51)
Born 1916			-0.102
			(0.22)
Born 1917			0.110
			(0.24)
Born 1918			0.123
			(0.20)
Born 1919			-0.414
			(1.00)
Constant	80.546***	83.056***	74.491***
	(236.59)	(29.71)	(49.63)
Observations	7209	1100	4789
Adjusted R-squared	0.00	0.03	0.00

Absolute value of t statistics in parentheses
* significant at 10%; ** significant at 5%; *** significant at 1%