"… Healthy, Wealthy, and Wise? Physical, Economic and Cognitive Effects of Early Life Conditions on Later Life Outcomes in the U.S., 1915-2005"

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

January 12, 2009

Abstract

Understanding the link between early and later circumstances is vital to enhancing our understanding of basic physiological, social, and economic mechanisms in operation over the entire life course, to identifying the protective factors that mitigate the negative effects of some early life experiences, and to designing effective interventions that reduce the long-term costs of adverse early life conditions. This project assesses how circumstances very early in life (e.g. economic privation, social isolation, proximity to environmental hazards or to medical care, exposure to unfavorable local disease environments) contribute to outcomes that are observed only decades later. It examines nationally representative data that follows several million individuals (who were born in the U.S. between 1895 and 1929 and who died in the U.S. between 1965 and 2006) from under age 5 until their death, with a wide range of information on early-life circumstances (at the individual, household, and community levels) and later life physiological, cognitive, and economic outcomes. This makes feasible for the first time linking early and later life circumstances across the twentieth century for a variety of sub-populations defined either in terms of demography (gender, race, specific birth cohort) or geography (region, city, neighborhood, block). The early-life information available for individuals includes exact place and date of birth, ethnicity, birth order, school attendance, the socio-economic status, ethnicity, and employment of both parents and neighbors, proximity to schools, stores, churches, and environmental hazards, and measures of the local disease environment. The later-life information includes adult height and weight, intelligence, educational and occupational attainment, income, disability, longevity, and specific cause of death.

A. INTRODUCTION

The impact of early-life circumstances on late-life outcomes has been a subject of intense interest in medicine and epidemiology since the work of Barker and others in the 1990s. This project evidence with which the impact on health late in life of conditions experienced by the individual very early in life (or even before birth) can be examined. In this respect, it provides a valuable input into the exploration of physiological processes that operate across the entire life course. Considerably less attention has been paid to the link between economic outcomes later in life and early life conditions, largely as a result of the lack of data appropriate for the investigation of these links. For example, most modern longitudinal surveys follow individuals for no more than 20 years and the longest-running (the Panel Study of Income Dynamics) began only in 1968 and so cannot yet follow individuals born into it past age 40. This project, in contrast, can follow individuals for more than nine decades. Together, the links between early conditions and later health and economic outcomes will enhance our ability to assess the impact of interventions to ameliorate the effects of adverse early life conditions.

The project generates the first body of data following individuals in the U.S. from under age 5 until death that (1) is sufficiently large to permit the analysis of the link between early and later life circumstances for even small populations defined by specific demographic or geographic characteristics, (2) includes both a wide range of information on the individual, household and community early in life and a wide range of later life outcomes, and (3) circumvents the biases generated by individual's decisions whether to continue their participation in longitudinal surveys by instead using a series of sources that document individuals' experiences at discrete points in time that can be linked across time to provide a longitudinal perspective.

Specifically, the aims of this project are to:

(1) Generate a new body of data for 2 million individuals who were born in the U.S. between 1885 and 1929 and who have died in the U.S. since 1965 with detailed information on both early-life conditions at the individual, household, and neighborhood levels and late-life physiological, cognitive, and economic conditions, by linking individual records from Social Security files, state death records, U.S. Census records, U.S. Army enlistment records, detailed plat and fire insurance maps, published mortality statistics, and records from Medicare and the Department of Veterans Affairs.

(2) Assess the biases in the data generated by linking these sources, and generate weights to ensure that the resulting population is nationally representative.

(3) Analyze the link between early-life conditions at the individual, household, and neighborhood levels and outcomes (IQ, height, weight, BMI, and education and occupational attainment) that can be observed at approximately age 25 (upon entry into the U.S. Army).

(4) Analyze links between early-life conditions and later-life outcomes that have already been suggested in other research (e.g. between exposure to infectious disease *in utero* or in infancy and later health; between household socio-economic status under age 5 and levels of income, disability, educational and occupational attainment, and chronic health conditions after middle age) but that have not been examined for large populations as they pass middle age and enter older ages .

(5) Assess differences across sub-populations (defined by demographic, economic, or geographic characteristics – race, gender, ethnicity, socio-economic class, location) in the link between early conditions and later outcomes.

(6) Explore the role of parents' (particularly mothers') backgrounds in the link between early

conditions and later outcomes, and assess the feasibility of integrating information from grandparents as well.

B. BACKGROUND AND SIGNIFICANCE

Both health scientists and economists have devoted considerable attention in recent years to the impact of early-life conditions on later-life outcomes. They differ, however, in the outcomes on which they focus. Health scientists have called attention to the impact of fetal and early childhood growth and environmental conditions on the later-life onset of chronic conditions. Economists have instead examined the effect of circumstances in the household and surrounding area on disability, intelligence, labor market participation, and income later in life; to the extent that there has been attention paid to specific conditions, they have been of interest largely insofar as they have an impact on these particular outcomes.

The observation that later outcomes can reflect early circumstances has a long pedigree in the health sciences. Writing in 1829, Villermé asserted that "the circumstances which accompany poverty delay the age at which complete stature is reached and stunt adult height." (Villermé, 1829; quoted in Smith and Ebrahim, 2001) The link between early nutrition (including nutrition *in utero*) and later health was initially examined in the context of the "experiment" generated by the Dutch famine of 1944-45. (Stein *et al.*, 1975) The "fetal origins of adult disease" hypothesis (Barker, 1992) offered a specific mechanism linking insults suffered *in utero* (such as poor maternal nutrition or exposure to infectious disease) to chronic health impairments much later in life, such as coronary heart disease (CHD), type 2 diabetes, and osteoperosis. Long-lasting effects on later health of insults suffered in childhood can also be identified (Eriksson *et al.*, 2003; Singhal *et al.*, 2002). As a result of the recognition that circumstances both before birth and through early childhood can influence later health outcomes, some have suggested abandoning the term "fetal origins of adult disease" in favor of "developmental origins of health and disability," which encompasses the entire sequence from prenatal through neonatal and infancy through early childhood (Gluckman and Hanson, 2006).

The epidemiological evidence on the "developmental origins" hypothesis for specific chronic conditions is substantial (Godfrey, 2006). The link between low birthweight and CHD in adulthood initially identified by Osmond *et al.* (1993) and Barker (1998) in data from the U.K. has now been found in a variety of other populations (Frankel *et al.*, 1996; Rich-Edwards *et al.*, 1997). The focus in recent years has shifted from low birthweight itself to more comprehensive measures of fetal growth, such as head circumference and shortness or thinness (Barker *et al.*, 1993; Martyn *et al.*, 1996; Forsen *et al.*, 1997; Barker, 1998; Eriksson *et al.*, 1999, 2001). At the same time, links between growth after birth and CHD have now been established (Osmond *et al.*, 1993; Barker, 1998; Eriksson *et al.*, 2001). A wide range of other later-life complications are now recognized as linked to growth *in utero* and in childhood: type 2 diabetes (Lithell *et al.,* 1996), stroke (Martyn *et al.*, 1996), hypertension (Huxley *et al.*, 2000), muskuloskeletal health (Harvey and Cooper, 2004), respiratory health (Barker *et al.*, 1991), cognitive function (Gale *et al.*, 2003 and 2004), and mental health (Susser *et al.*, 1996; Sacker *et al.*, 1995). In a summary of recent research on the mechanisms linking early conditions and later outcomes, Gluckman *et al.* (2008, p. 68-69) suggest that circumstances even prior to conception – as far back as a generation or more – might also be important determinants of an individual's leter health: "The developmental cue is not limited to the nutritional environment during the period of gestation; rather, the information passed to the fetus or neonate from conception to weaning is a summation of maternal nutritional experience, integrating a lifetime of signals from the mother and perhaps even the grandmother."

Doblhammer and Vaupel (2001) have identified a specific characteristic that is easily observed (quarter of birth) and plausibly related to intrauterine growth, and found that its effect on longevity for those age 50 and over is pronounced: in the Northern Hemisphere, those born in the fourth quarter live longer than those born in the second quarter, while this pattern is reversed in the Southern Hemisphere. Second quarter births in the Northern Hemisphere and fourth quarter births in the Southern Hemisphere both follow winter months during which fetal growth maybe impaired by poor maternal nutrition. Costa and Lahey (2005) find that the effect of season of birth was considerably greater in a population that was born in the first half of the nineteenth century.

Recent research has moved beyond the examination of fetal, neonatal, and early childhood growth as a contributor to later adverse outcomes and focused instead on aspects of the physical environment. Exposure to lead early in life has been identified (through brain imaging) as a source of changes in the physical structure of the anterior cingulate cortex, which is "consistent with and potentially explanatory for cognitive and behavioral problems previously associated with lead exposure." (Cecil *et al.*, 2008, p. 744) Similarly, exposure to cadmium in childhood is associated with reduced kidney function in adulthood to a greater extent than exposure in adulthood. (Trzcinka-Ochocka et al., 2004) The effects of early growth and environmental influences are not independent: a particular growth trajectory may increase susceptibility to environmental influences. (Gluckman *et al.*, 2008, p. 62)

Like epidemiologists, economists have recognized for some time now that later-life economic outcomes can be shaped by both circumstances early in life over which parent and child exercise no control ("endowments") and the choices made subsequently by both (Becker and Tomes, 1976). In recent work by Heckman (2008), these issues are addressed explicitly in a framework that allows early investments in an individual's capacities to enhance the productivity of later investments and that attempts to identify the optimal time for positive interventions or the most harmful time for individuals to experience negative shocks to their capacities. In a survey of the work on early childhood conditions and later-life economic outcomes, Currie and Madrian (1999, p. 3352) state that "studies suggest that health in childhood could be an important determinant of future labor market success, a question that has received little attention to date, perhaps because of data limitations."

Some successes have been achieved in linking early life experiences and later economic outcomes since Currie and Madrian's 1999 assessment, many of which are summarized by Currie (2008). Case, Fertig, and Paxson (2005), using data from Britain, find that individuals who experienced poor health in childhood were less educated, less healthy, and lower in social status as adults than individuals whose childhood health was better. Black *et al.* (2007), using data from Norway, show that low birthweight is associated with lower IQ, educational attainment, and earnings in adulthood. Johnson and Schoeni (2007) also find that low birthweight leads to lower IQ, educational attainment, and earnings, using data for the U.S. from the PSID. Using 4,562 males (2,693 of whom were decedents) from the National Longitudinal Survey of Older Men (NLS), Hayward and Gorman (2004) examined mortality at older ages as a function of early-life characteristics (though "early life" in this survey refers to circumstances when the respondent was age 15), finding that "men's mortality is associated with an array of childhood conditions, including socioeconomic status, family living arrangements, mother's work status, rural residence, and parents' nativity." (p. 87)

Like epidemiologists, economists have also considered the impact of the physical environment. The contagious disease environment experienced in childhood is associated with worse health in young adulthood: individuals who lived in an environment characterized by greater morbidity were at greater risk for metabolic syndrom by age 35. (Margolis, 2008) Feyrer *et al.* (2007) show that cognitive functioning is better for individuals who lived in regions characterized by iodine deficiency after the introduction of salt iodization. Almond (2006) shows that the impacts later in life for individuals who were *in utero* at the time of the 1918-19 influenza pandemic were worse than the effects for those who had been born before the pandemic or who were born later but were not *in utero* when the pandemic occurred: higher rates of disability, lower educational attainment, and lower incomes as late as the 1980 census, more than 60 years after the exposure to influenza. The Feyrer *et al.* and Almond studies are noteworthy because they both attempt to overcome the problem of omitted variable bias (in Almond's case, through the examination of a short, sharp treatment – the pandemic lasted only 9 months – and in Feyrer *et al.*'s case through the examination of a treatment – salt iodization – that also occurred quickly and should have had an effect only in those areas where iodine was previously deficient). Reyes (2007) has examined the impact of lead exposure early in life and shown that it leads to increased adult criminal activity.

The present project generates a body of data containing roughly 2 million individuals who are followed from their birth (in the years 1885-1929) to their death (in the years 1965-2006), with detailed information on their circumstances early in life, their condition at around age 25, and their later life income, disability, health, and longevity. This data can overcome many of the shortcomings of the data employed in previous studies. Assessing the strength of the links between conditions either *in utero* or in childhood and later health and economic outcomes (particularly after middle age) has been hampered by a lack of data following individuals from birth until much later in life. This problem is particularly acute for the U.S. which lacks a centralized system of vital registration, making it difficult until now to link individuals from records documenting their birth and early life to later life events such as marriages, census enumerations, military conscription records, government pension records, and death records. As a result, there are no large-scale studies for the U.S. that follow individuals across their entire lives. Though several British studies such as the 1958 National Child Development Study (Case, Fertig, and Paxson, 2005) do indeed follow individuals over a substantial portion of their lives (from birth to age 42 in the NCDS), these studies are generally small in size, subject to attrition bias or retrospectivity bias (if the information on early conditions is collected only after entry into the survey), and lacking in detailed information on the individual's physical environment early in life.

The most significant respect in which this project represents an improvement, then, is in its ability to cover the life course from beginning to end. If early conditions have an effect on disability at older ages (after middle age), or on longevity, or on specific causes of death, or on lifetime levels of income or educational or occupational attainment, these effects cannot be observed in existing data sources. To the extent that the sources that have been employed to date in assessing the later impact of early conditions have paid so little attention to life later than "middle age," researchers have perhaps missed a substantial part of the impact. The data generated in this project will make it possible to identify these impacts that occur much later in life, right up to and including impacts on longevity and specific cause of death.

Another respect in which this project improves our ability to understand the link between early and later conditions is its inclusion of detailed contextual data early in each individual's life: it includes the characteristics of neighbors and neighborhoods, including proximity to both environmental hazards (e.g. sources of airborne lead pollution, unsanitary water, prevalence of contagious disease) and resources that could mitigate some of the negative effects of poor early-life conditions (schools, churches, parks, retail establishments). A shortcoming of existing data is the absence of detailed information (other than perhaps place of birth) on the local physical environment with which an individual had contact early in life.

A final important factor that distinguishes this project from previous efforts to assess the link between early and later conditions is its sheer size. It has a sufficiently large number of observations that even small sub-populations (classified by race, gender, birth cohort, ethnicity, or location) can be analyzed, and the range of variables that can be analyzed is broad enough (including variables such as IQ test scores at enlistment in the U.S. Army that have never been used before at the individual level) to allow the analysis of a number of pathways leading from early conditions to later outcomes.

In both its long time horizon (following individuals from birth to death) and its wealth of contextual information, the present project is similar to the Early Indicators Project at the University of Chicago's Center for Population Economics examining the experiences of Union Army veterans as they aged. There are two important differences, though: (1) the present project includes both men and women, and individuals who served in the U.S. military and those who did not, so its coverage is more representative of the U.S. population; and (2) the present project includes deaths occurring as recently as 2006, and examines the experiences of cohorts still moving through the retirement and health care systems today. The present project is also similar to the National Longitudinal Mortality Study (NLMS) which has linked Current Population Survey and other records since the early 1970s to death records. The present project differs from the NLMS in two ways: (1) the NLMS is projected to have 500,000 deaths by 2009, only a quarter of the total in the present project; and (2) though the NLMS is in the process of incorporating tract-level neighborhood contextual information (including exposure to pollution), this information will be available only at the time individuals enter the sample (through the CPS or the 1980 census), providing "childhood" context only for individuals who entered the sample at young ages, so the deaths for which early-life neighborhood context can be examined will occur no later than age 40 (individuals who were age 5 in 1973). The present project is also mechanically similar to the study of extreme longevity among African-Americans by Preston *et al.* (1998) which linked individuals who were age 85+ in January, 1985 from death certificates back to their appearance in the 1900 and 1910 U.S. population censuses. The present project is considerably larger than the 582 observations used by Preston *et al.*, includes individuals located in 4 U.S. censuses (1900-1930) all of which have now been completely indexed and can be electronically searched, with a large fraction linked to more than one census, includes all racial groups, and includes deaths at ages as young as 35 (for individuals who were born in 1929 and appeared in the 1930 census and who died in 1965 at the beginning of the computerized Social Security file the present project employs).

The project contributes to our understanding of the process of aging in three respects: (1) providing new data to analyze the link between early conditions and later outcomes, making it possible to uncover specific biological and social mechanisms that operate over the entire life course, from birth to death, and even across generations; (2) providing a life-course perspective on health and disability that will be useful in planning for the needs of future generations as they entire the second half of their lives, by assessing how their needs evolve as the circumstances through which they passed in early life have evolved; and (3) providing information vital for an accurate assessment of the costs and benefits of interventions (e.g. improvements in maternal health and public sanitation) that may have a substantial payoff only decades after the intervention is conducted, which may be particularly relevant for countries still undergoing the process of economic development.

C. RESEARCH DESIGN AND METHODS

The primary aim of the project, generating new data on 2 million individuals followed from early life to death, consists of 3 broad tasks: (1) substantially expanding the linkages we have already investigated (among the 1900-30 U.S. censuses, Social Security's Death Master File, and the U.S. Army's enlistment records), (2) using data from Social Security's NUMIDENT file to verify matches; (3) and adding Social Security data on disability and income, published mortality data, Medicare and Veterans Department data on interactions with those systems, state death records, and plat and fire insurance maps that describe the physical environment within which the individual lived early in life. The secondary aims, to analyze several links between early and later conditions already suggested in the literature and to develop new analyses that exploit the unique features of the new data the project creates, will depend on the successful completion of the primary aim. As a result, the bulk of the attention here will be devoted to the data collection.

1. Generating the Data

In its simplest form, the data collection we undertake consists simply of locating the same individuals in two sources (the U.S. Census of Population and the Social Security Administration's Death Master File). The census data itself provide important early life information, but because it describes the exact place of residence at the time of the census (including the street address in cities and towns), the census data also makes it possible to attach additional contextual information (such as characteristics of neighbors, mortality statistics for the place of residence, and the locations of nearby community assets and hazards). The Death Master File likewise provides useful information by itself (date and place of death), but because it provides a Social Security number for each decedent, it will also make it possible to attach additional information to each linked record (such as other Social Security information in the NUMIDENT file that will make it possible to verify matches, Social Security pension and disability information, and Medicare and Veterans Department data).[1] Finally, linkage of U.S. Army enlistment data to either source makes it possible to analyze height, weight, occupational and educational attainment, and IQ (all measured around age 25) as either a result of earlier conditions (if the linkage is to a prior census) or as a predictor of subsequent outcomes (if the linkage is to the DMF and the sources to which DMF records can be linked). Linkage can be done either to the large public use samples for the 1900-30 censuses (Ruggles et al., 2008) or to the entire collection of digitized and indexed 1900-30 census manuscripts available on-line. Attention will be limited for censuses only as recent as 1930 because only these censuses have no restrictions on the release of identifying information.

The first complication that arises to this simple plan is that the DMF has more complete coverage of deaths for ages 65 and above than for deaths below age 65 (since the file is generated by reports of deaths that are required to stop Social Security payments). This is shown in Figure 1 by the higher rates of reported deaths in the DMF for ages 65 and over, particularly before the 1970s. This means that the linked data can examine late-life outcomes only for individuals who survive to age 65. In order for an individual to be linked to the DMF,

[1] An alternative source of death data, the National Death Index (NDI), has two distinct disadvantages compared to the DMF: the NDI only begins in 1979, and the cost of obtaining large numbers of records is prohibitive. We are presently working with a publicly-available version of the DMF, the Social Security Death Index (SSDI) which contains each decedent's full name, date of birth, date and place of death, and Social Security Number.

then, that individual must survive to age 65. A second complication is that the system through which Social Security obtains death information changed after 1965, making it easier for SSA to obtain death information from this date. This is shown in Figure 1 by the sharp increase 1965-68 in the reported deaths in the DMF even for deaths at age 65 and over. This means that death data will be available mainly for those individuals who survived to 1965. Because Social Security will release the other information in its files (identifying information from the NUMIDENT, pension benefit levels, reports of disability payments) only for individuals whose deaths can be verified in the DMF, and a Social Security number from the DMF will be necessary to add Medicare and Veterans Department data, the linkage process will be constrained in two respects: it will allow examination of outcomes only for individuals (1) who died at age 65 and later, and (2) who died in calendar year 1965 and later.

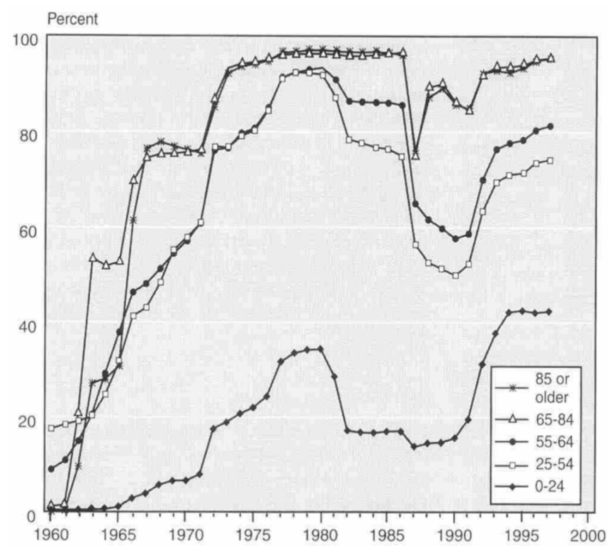

FIGURE 1. Coverage of Deaths in SSA's Death Master File by Age, 1960-2000. Source: Hill and Rosenwaike (2001/2002).

The project's research design relies on two fundamental elements: the Social Security Administration's Death Master File (DMF), derived from the NUMIDENT file maintained by SSA (Panis *et al.*, 2000) and the 1900-1930 U.S. population censuses. Once individuals are linked between 1900-30 census records and the DMF, it is immediately possible to assess the link between early-life conditions and one late-life outcome: longevity.

The NUMIDENT file provides information from the individual's original SS-5 form filed at entry into the Social Security system (date and detailed place of birth, full names for both parents, including mother's maiden name), as well as information on the individual's date of death. Use of the NUMIDENT and the DMF based on it raises two concerns. The first is that there were broad classes of workers (domestic workers, farm workers, the self-employed) who were not originally covered under Social Security, so SSA records may miss a substantial fraction of the population. The 1940 U.S. Census included a unique question on whether the respondent had received a Social Security card. Figure 2 shows the distribution of positive responses for two farm states, by age in 1940 and migration status. Even as early as 1940, between 60 and 70 percent of males born in 1920 and earlier had already entered the Social Security System.

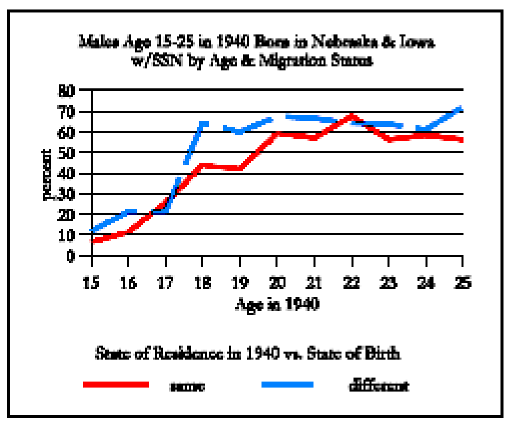

FIGURE 2. Iowa and Nebraska Males with SSN in 1940 by Age & Migration.

The NUMIDENT file is not essential if unique matching to the DMF can be done without the additional identifying information. In fact, in each birth year from 1885 to 1929, from 74 to 85 percent of all individuals in the DMF have a unique combination of surname, given name, middle initial, and year of birth; if month of birth is examined as well, from 92 to 96 percent of all individuals who were born in these years and whose deaths appeared in the DMF were unique. This makes it possible to match virtually all male births from 1885-1899 to the DMF if the deaths occurred at age 65 or later and occurred in year 1965 or later (as the 1900 census records both month and year of birth for each individual), and from 74 to 85 percent of individuals for whom we can infer only year of birth at the census date and 92 to 96 percent of individuals for whom we can infer both month and year of birth at the census date (under age 2 in 1910 and under age 5 in 1920 and 1930; for these ages, enumerators were instructed to report completed age as of the census reference date in full years and months). But for matching women from the DMF to the census, the NUMIDENT information (particularly on parents' names) is essential.

The use of the NUMIDENT file to ascertain the place of birth and parents' names introduces an additional complication: this file was computerized only in the mid-1970s. If individuals filed claims for pension or disability benefits before their files were computerized, the detailed identifying information from the original SS-5 forms was removed from the file and sent to the local Social Security offices administering the claims. This information was then not present in the file when it was computerized, so as Elo *et al.* (2004) note, "information on place of birth, race/ethnicity, and maiden name is often missing for persons who filed for social security benefits or enrolled in Medicare before the application file was computerized in the mid-1970s." (p. 115) Figure 3 shows the percentage of observations in the NUMIDENT file with the original SS-5 materials still in place at the time the file was computerized. It rises rapidly from births around 1910 to births around 1915. The 20 percent of pre-1910 births with complete SS-5 information in the NUMIDENT are the result of individuals who entered the system but never claimed benefits, most likely because they died before achieving eligibility for a pension.

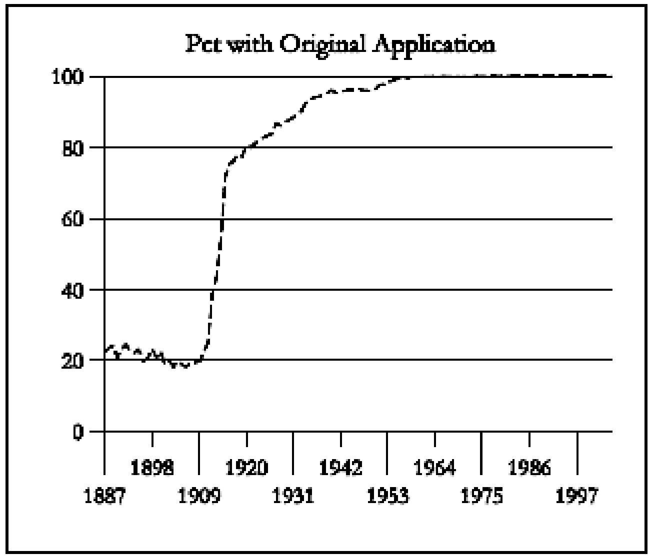

FIGURE 3. Entries in Social Security's NUMIDENT file with Original SS-5 Information Present, By Birth Year.

The basic linkage between Social Security records and the census is done in two ways: (1) using the public use samples (these are presently 1 percent, but we will have access to pre-release versions of the 6 percent 1900 sample and the 5 percent 1930 sample, which will be assumed in the following calculations); and (2) using the manuscripts themselves. We will link all individuals who were under age 45 in 1930, under age 35 in 1920, under age 25 in 1910 and under age 15 in 1900, for two reasons: this will yield a large number of linked siblings, making possible the use of household-level fixed effects in the data analysis; and it will make it possible to exploit multiple observations on the same individual, by linking individuals back to prior censuses and forward to subsequent censuses. For example, an individual who was age 44 in the 1930 census can be linked back to the 1920, 1910, and 1900 censuses, yielding information on circumstances at 4 dates. An individual who was age 4 in 1900 can be linked forward to the 1910, 1920, and 1930 censuses. Table 1 shows the range of information available for each linked individual, others in the individual's early-life household, and the individual's early-life neighbors.

Approach (1) can be done for individuals who survived to 1965 and who died at age 65 or later. These restrictions yield a predicted total linkage of 1,407,163 males. This assumes that individuals experience the survivorship for their birth cohorts observed in the 1960-200 censuses and the 2007 American Community Survey, and that 90 percent of pre-1900 births, 93.5 percent of 1908-09, 1916-19, and 1926-29 births, and 79 percent of all other births have exactly one match in the DMF. For females (for whom the additional restriction that they appear in the NUMIDENT with their original SS-5 information must be satisfied), a predicted linkage of 1,296,562 individuals results. This imposes the same restrictions on survivorship but assumes that the percentage appearing with original SS-5 information in the NUMIDENT file is zero for pre-1915 births, then rises from 72 percent for 1915 births to 87 percent for 1929 births. A total of 2.7 million individuals will be linked in this manner. For each individual, information on the demographic characteristics of the neighborhood (race, age structure, ethnicity, population density) can be retrieved electronically from the on-line census indexes.

	1900	1910	1920	1930	DMF	NUM	MBR	DIS
Name	X	X	X	X	X	X	X	X
SSN					X	X	X	X
Residence at Census								
State, County, City	X	X	X	X				
Street Address	X	X	X	X				
Father's Name	X	X	X	X		X		
Mother's Name	X	X	X	X		X		
Birthplace								
State	X	X	X	X		X		
City/County						X		
Date of Birth					X	X		
Date of Death					X	X		
Residence at Death					X	X		
Age	X	X	X	X				
Sex	X	X	X	X		X		
Race	X	X	X	X				
Occupation	X	X	X	X				
Industry	X	X	X	X				
Unemployment		X		X				
School Attendance	X	X	X	X				
Home Ownership	X	X	X	X				
Home Value				X				
Monthly Pension							X	
Disability		X						X

Note: DMF=Death Master File, NUM=NUMIDENT file, MBR=Master Beneficiary Record File, DIS=831 Disability File (see Panis *et al.*, 2000, for details).

TABLE 1. Variables in Census and Social Security Files.

For some analyses, however, even the high-density 1900 and 1930 samples are inadequate. For example, we will analyze the impact on residents of Omaha, Nebraska of proximity to a lead smelter on the city's eastern edge that operated from the 1870 to 1997 and is the subject of an EPA priority clean-up today. Using the electronic indexes, we will retrieve all of the residents of Omaha in the 1900-30 censuses who lived within 5 miles of the site. Similar geographically-focused linkages will be performed for the project on the impact of lead water pipes, on the impact of parents' deaths in industrial accidents in mining communities, and on the impact of the reversal of the flow of the Chicago River. These additional linkages will add several hundred thousand observations to the linked data.

Both this geographically-focused linkage and the searches in other 1900-30 census manuscripts for individuals drawn from one of the 1900-30 public use samples are done with a simple Perl script that is run on several personal computer simultaneously. The script executes a search in the on-line census index maintained by Ancestry.com and retrieves the HTML code behind the on-screen information that the search generates. These HTML files are converted to flat text and then parsed with a high-level language (SAS) that extracts all of the information that was transcribed from the original schedules in the indexing process. It is then necessary to transcribe from the original schedule the information not extracted in the indexing process (e.g. occupation, home ownership, education, exact street address). Arrangements have been made with a data transcription company to process up to 100,000 such records.

Once the linkage is done between the census (either the public use files or the manuscripts) and the DMF, the addition of other sources that contain SSN as an identifier is straightforward. The Medicare data on physician visits, hospitalizations, drug and durable medical equipment purchases, and hospice care are available from 1991 to 2006 with a Medicare identifier that is unique for each recipient and keyed to the recipient's SSN (in the MedPAR and Denominator files). These files with identifiers are available through the Research Assistance Data Center for a fee of roughly $900 per year of data for each file. In the same way, the Veterans Department Patient Treatment File (for in-patient care) and Outpatient Care File (1976 to present in many cases) have identifiers that are derived from SSNs; the original SSN can be obtained for research purposes with permission from the department. Examples of the information contained in these files are shown in Table 2.

	Medicare	Veterans Department
Age	X	X
Sex	X	X
Race	X	X
Residence	X	X
Admission Date	X	X
Discharge Date	X	X
Days in ICU	X	
Days in Coronary Care	X	
Diagnosis	X	X
Surgical Procedure	X	X
Outpatient Services	X	X
Durable Medical Equipment	X	

TABLE 2. Variables Available in Medicare (1991-present) and Veterans Department (1976-present) Restricted Files.

The next source to which individuals can be linked is state death records. All states now provide electronic versions of their death certificate transcriptions to the National Center for Health Statistics on an annual basis. These certificates are used to construct NCHS's Mortality Detail File. There is insufficient identifying information in the files available through NCHS, however, to make them useful for linkage purposes. As a result, in order to obtain cause of death information, it is necessary to acquire each state's death certificates in electronic form. The standard U.S. death certificate in use by the early 1970s provides most of the necessary identifying information, as well as ICD codes for primary and contributing causes of death. Many states also include additional information useful in the linkage process (SSN, full name of decedent and parents, detailed place of birth), as well as information on occupation, industry, and final educational attainment. Table 3 shows the contents of the standard death certificate in use at various dates. The only items that are not consistently available throughout the period from the early 1970s to the present are education and city or town at birth (both introduced with the 1989 revision). Though some states have at times departed from these standards (for example, Michigan began to report birthplace only in 1979 and SSN only in 1989), death records nonetheless represent a crucial additional source of information.

The principal shortcoming of this source (apart from the complexity of obtaining records from 51 separate bureaucracies) is the relatively short time-frame they cover: the DMF begins in 1965, but most computerized state death records only begin in 1970 or 1972. We have obtained additional death records for California that extend back to 1945, although these records do not include cause of death. We have already used these records to assess how much of the impact of early conditions on later outcomes is missed by a focus only on deaths at older ages.

We have already obtained death records for four states (California, Massachusetts, Michigan, and Missouri) and have identified the relevant contact person in the vital statistics section in each of the remaining states and the District of Columbia. We have budgeted funds for acquiring all of these data from 1970 to the present. To the extent that many of these records include the surnames of the decedent's parents, this source represents an additional avenue through which women can be included in the analysis: even if they were born before 1915 (and therefore were unlikely to have their original SS-5 information in their NUMIDENT record at SSA), they would still enter a state death record is they died at any date since roughly 1970. The inclusion of mother's maiden name also makes it possible (after a decedent has been linked to a census) to extend the set of "early-life" conditions back a generation, by locating mothers in the households in which they themselves lived as children (see "6. Explore the Role of Parents' (Particularly Mothers') Backgrounds" below).

	1968	1978	1989	2003
Name	X	X	X	X
SSN	X	X	X	X
Sex	X	X	X	X
Race	X	X	X	X
Date of Birth	X	X	X	X
Birthplace				
State	X	X	X	X
City			X	X
Usual Occupation	X	X	X	X
Industry	X	X	X	X
Education			X	X
Father's Name	X	X	X	X
Mother's Maiden Name	X	X	X	X
Residence				
City/Town	X	X	X	X
Street Address	X	X	X	X
Cause of Death				
Immediate Cause	X	X	X	X
Other Conditions	X	X	X	X

TABLE 3. Contents of Standard U.S. Death Certificate, 1968-2003.

The U.S. Army's Serial Number Electronic File 1938-46 that we have obtained from the National Archives provides additional information on male decedents at approximately age 25. This file records the information collected at the induction site for men who had been found fit for service by a local draft board physician (Edwards and Hellman, 1944). These records document only those deemed fit for service and thus exclude individuals with physical or mental characteristics that disqualified them for military

	July 1940 to February 1943	March 1943 to June 1943	June 1943 to March 1944
Name	X	X	X
Birth Year	X	X	X
Birth State	X	X	X
Race	X	X	X
Residence County	X	X	X
Occupation	X	X	X
Marital Status	X	X	X
Family Size	X	X	X
Height	X		
Weight	X		
AGCT Score (IQ)		X	
Branch Assignment	X	X	X
Observations (millions)	5.2	0.5	0.1

TABLE 4. Information in U.S. Army Serial number Electronic File by Enlistment Date.

duty. The standards for these characteristics evolved over the course of the war as manpower needs changed, and we have identified those changes, making it possible to assess how the selection bias in this data changed over the course of the war. This file is the only significant source of information on the population of males inducted into the U.S. military prior to 1960, as more than 80 percent of the individual military personnel files held by the National Archives at the National Personnel Records Center in St. Louis were destroyed by a fire in July, 1973. As we have shown above, we have discovered (see section "4. The Impact of Lead" in "C. Preliminary Studies") that the enlistment file contains not just height, weight, and other characteristics at around age 25, but also contains the results of the Army General Classification Test, an intelligence examination, for more than half a million inductees who entered the Army from March through mid-June of 1943.

Linking these records to pre-1940 census records and to later-life sources such as Social Security records is straightforward. Table 4 shows the characteristics available in the file, both for purposes of identification and matching and for purposes of analysis. In each year, full name, state of birth, and year of birth are provided. We have already linked a sample of these records back to the 1920 census and forward to California death records. We will link all of the records with AGCT scores back to the 1920 and 1930 population census and forward to the DMF and other Social Security sources, as well as to state death records and Veterans Department records. For the records we link between census public use samples and Social Security records, we will identify any individuals who served in the U.S. Army in the enlistment file as well; most of those linked in this way will contain height and weight rather than IQ.

Another source to which individuals can be linked based on their locations (exact street address in cities and towns; township and enumeration district in rural places) at the time of pre-1940 censuses is the large collection of highly-detailed plat maps and fire insurance maps that are now available in digital form. Figure 4 shows an example of a plat map for part of Otter Creek Township in Jersey County, Illinois (1916). The maps identifies the locations of all land owners, along with features of the physical landscape such as rivers. These maps make it possible to assess the land acreage of residents in rural places, as well as their proximity to other specific individuals. These maps also show the locations of schools and churches. In urban places, fire insurance maps like those created by the Sanborn Map Company serve the same purpose. These maps exist for most major cities and towns and cover the period from the 1880s through the 1950s. The level of detail is sufficient (the scale is 1 inch=50 feet; for an example, see Figure 5) to identify the dimensions and building materials of individual structures at each street address, as well as the proximity to water and sewer pipes, railroad and streetcar lines, retail establishments, schools, churches, hospitals, bodies of water, parks, and specific factories.

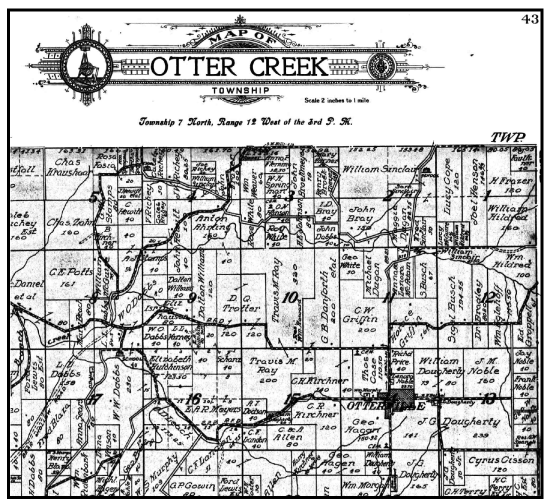

FIGURE 4. 1916 Plat Map for Part of Otter Creek Township, Jersey County, Illinois.

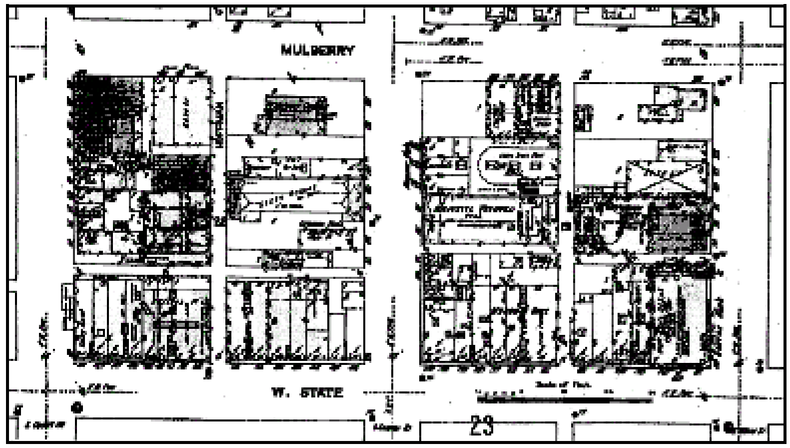

FIGURE 5. Sanborn Map for Part of City of Rockford, Winnebago County, Illinois, 1913.

More than 1,200 digitized county plat maps are available on-line for the period 1864-1918, while Sanborn maps are available for more than 12,000 cities and towns from 1867 to 1970. For both sources, the challenge will be georeferencing the maps in a way that allows the analysis of neighborhood characteristics. At this time, we anticipate that entire maps will be georeferenced only for those locations for which we draw geographically-focused samples of individuals from census manuscripts (e.g. Omaha, and parts of Chicago and St. Louis). For other places, we will instead identify and georeference a subset of the total information available on these maps (e.g. sites identified as environmental hazards such as factories and local community assets that could reduce the negative impact at later ages of some early life conditions such as low SES or crowding – parks, schools, churches, hospitals, and retail establishments).

The final source to which individuals will be linked will be published historical death data from annual U.S. Census Bureau Mortality Statistics Reports. These volumes extend from 1890 to 1936 and report deaths by month, cause, and age for locations as small as medium-size cities and towns. We have included in our budget funds sufficient to transcribe these statistics for the period 1890-1930. This will make it possible to assess how the early-life disease environment influenced later-life outcomes. These will be supplemented with tabulations of deaths in specific cities by neighborhood or ward that have been collected as part of the Early Indicators Project (for example, we now have in hand data generated by the health departments in Boston, Detroit, St. Louis, and San Francisco)

At the completion of the linkage process, we anticipate that we will have 2 million individuals linked from their early-life household (described in either census public use samples or extracts we create from census manuscript schedules for entire communities or geographic areas) to their death date (recorded in Social Security's Death Master File). Most of these will also be linked to additional Social Security records (their monthly pension benefit and any disability they claimed) and to state death records (that include cause of death). Roughly 1/3 will be identified in Medicare records, and 1/4 of males will be identified in Veterans Department health system records. Half a million records will be linked from U.S. Army enlistment records with IQ test results back to their early-life household, and another 200,000 of the males linked from census public use samples (half of all males born 1915-1926 who survived to 1990 served in the U.S. Army in World War Two) to later-life sources will be linked to U.S. Army enlistment records with height and weight reported at enlistment. All individuals linked to the Army records will also have information on residence, occupation, and family structure at the time of enlistment (around age 25). All individuals linked to 1900-30 censuses will also be linked to published mortality statistics for their place of residence (county or city/town). Finally, individuals in a subset of locations (roughly 100,000 individuals in all) will be linked to detailed maps of the places where they lived early in life, providing information on proximity to neighbors, community assets, and environmental hazards.

D. PRELIMINARY ANALYIS

The 1900-30 U.S. population censuses exist in two forms that can be linked to the DMF: (1) as manuscript schedules themselves which are fully indexed (by name, birth year, state of birth, residence at the time of the census, and names of both parents), partially transcribed, and available as images on-line; and (2) as 1 percent Public Use Files generated through the IPUMS project (Ruggles *et al.*, 2008) – the 1900 sample will be expanded to 6 percent in 2009 and the 1930 sample will be expanded to 5 percent in 2012. The identifying information available in the census (manuscripts or IPUMS) that can be used to link individuals to the SSDI is full name and year and month of birth (exact year and month of birth are available directly only in 1900; it can be inferred from the age in years and months reported for individuals under age 2 in 1910 and under age 5 in 1920 and 1930, together with the census reference date).

Linkage between the SSDI and the census manuscripts can be done by searching the on-line index of the complete census for a particular year, while linkage between the SSDI and the IPUMS can be done with a simple matching algorithm. The advantage linking between the census manuscripts and the SSDI is that it is [almost] infinitely scalable (we can link as many records as we draw from the SSDI, so a final linked population of several million is feasible,) so we can look at very small populations defined either in terms of demography (race, ethnicity) or geography (we will examine the entire population of Omaha, NE and assess the impact of proximity to a lead smelter in operation from the 1870s to 1997 located in what is now the city's downtown; this remains an EPA cleanup priority site). The advantage of linking between the IPUMS and the SSDI is that it provides broad geographic coverage and can be done at relatively low cost, as it involves little additional data transcription; the disadvantage is that even 5% samples will be inadequate for examining some neighborhood effects and the total resulting linked population is limited by the size of the original public use sample from which the linkage begins.

In both cases, a tolerance is allowed in the spelling of surnames (to allow for the mistranscription of names by census enumerators) and in the month of birth (one month on either side of the month predicted on the basis of the reported age in years and months in the census, to allow for the possibility of rounding error). Because of the limited identifying information available in the SSDI, linkage from the SSDI to either the census manuscripts or the IPUMS results in numerous cases in which one SSDI case was matched to multiple census cases. We are in the process of obtaining 400,000 records from the NUMIDENT file (with full place of birth and the names of both parents) that will allow us to resolve these ambiguities. In the results reported below that entail linkage involving the SSDI, we have used only observations that could be matched to only one individual, or in which it was possible to rank potential matches (by the closeness of the spelling of the surname or the proximity of the SSDI-reported month of birth to the census-reported month of birth) and choose the “best” match.

Though the present linked samples are biased toward those with unusual combinations of surname, given name, and month and year of birth, these individuals are not different in terms of their longevity (in the SSDI) from the general population; they are also not statistically distinguishable in terms of any observable characteristics (in the census manuscripts or IPUMS) from the general population. In any case, this bias will be eliminated when the NUMIDENT information is incorporated into the linkage process, as this source includes detailed place of birth as well as the full names of both parents. To see how thoroughly the NUMIDENT information eliminates ambiguity, consider linking an individual named “John Smith” from the SSDI to the 1920 census. The 1920 census manuscripts contain 30,683 individuals with this name. If we use the information on year of birth from the SSDI to confine our search to individuals born in 1918, we can narrow this down to 498 individuals. If we look only for individuals born in June, 1918 (whose reported completed age on January 1, 1920, the reference date for the 1920 census, would be 1 year and 6 months), the number of "hits" falls to 60. Finally, if we use information from the NUMIDENT on either place of birth (and assume that this is the same as location in the 1920 census) or the names of both parents, all ambiguity is resolved. For example, in the 1920 census there is exactly one "John Smith" born in June, 1918 and residing in New York City in Manhattan; there is exactly one "John Smith" born in June, 1918 whose parents are "John" and "Anna."

The linkage process just described begins with the SSDI which does not contain the maiden name for women who have married and changed their names. As a result, direct linkage to the SSDI can be done only for males. When we receive the NUMIDENT extract, however, we will be able to link woman as well, as the NUMIDENT contains the father's full name (from which a female child's maiden name can be inferred). As an alternative, California's birth and death records for women were matched for 1940-2004 deaths. This was done ignoring surnames entirely and using only given name, exact date of birth, and place of birth/death (CA) as linking criteria, resulting in 7,809 matches. Together with 15,328 males matched in the same records, this makes it possible to assess differences by gender in the impact of early conditions on longevity.

	(1)	(2)
Born January-March	-0.258	0.984*
	(1.19)	(1.83)
Born April-June	-0.623***	-1.540***
	(2.81)	(2.72)
Born July-September	-0.398*	0.213
	(1.85)	(0.37)
Household Mortality Rate	-0.713	-5.019***
	(1.06)	(2.77)
State Mortality Rate		-0.160
		(1.20)
Constant	80.546***	83.056***
	(236.59)	(29.71)
Observations	7,209	1,100
Adjusted R^2	0.00	0.03

Absolute value of *t*-statistics in parentheses. * p < 0.10 ** p<0.05 *** p<0.01

TABLE 5. OLS Regressions on Age at Death, Males Linked from 1900-10 IPUMS to SSDI Who Died Age 70-95.

The final linkage we have conducted in our preliminary work uses the U.S. Army's Serial Number Electronic File (1938-46). This file contains enlistment data for 9.2 million individuals who served in the Army, Enlisted Reserve Corps, and Women's Army Auxiliary Corps. Each record reports the individual's full name, year of birth, and state of birth, together with occupation and educational attainment, marital status, family size, height, and weight at the time of enlistment. We have also discovered (see Section C.4 below, "The Impact of Lead") that for a period of 16 weeks from early March, 1943 through the middle of June, 1943, the file does not contain the enlistee's weight; instead, the data punched into this field is the score on the enlistee's Army General Classification Test, essentially an IQ test that was the predecessor to the modern AFQT. These scores are reported for 523,956 individuals. These records can be linked back to the U.S. census using name, year of birth, and state of birth, and forward to the SSDI using name and year of birth.

1. Longevity, Season of Birth, and the Local Mortality Regime

The 1900 and 1910 IPUMS files were linked to the SSDI, resulting in 7,209 unambiguous matches. These linked records were used to assess the association between characteristics observed early in life (under age 5 in 1900, under age 2 in 1910) and longevity for individuals who died between the ages of 70 and 95 (this restriction was imposed in order to ensure that each individual was at risk to survive to all ages in the range; the SSDI records begin in 1965 and end in 2005). Table 5 shows OLS regression coefficients for several variables of interest (not shown are controls for region, city size, family size, and year of birth). The “household mortality rate” is constructed using two questions unique to the 1900 and 1910 census: mothers were asked to report the number of children ever born and the number of children surviving. The variable is 1 - (surviving/born) and ranges from 0 (all children born survived) to 1 (no surviving children); higher values indicate a higher-mortality household. The “state mortality rate” is the published state-level mortality rate in the individual’s year of birth.

Two conclusions emerge from this simple analysis: (1) individuals born in the second quarter of the year live between 0.6 and 1.5 years less than individuals born in the fourth quarter (the excluded category), consistent with the findings of Costa and Lahey (2005) and Doblhammer and Vaupel (2001); and (2) when the state-level mortality rate is included (column 2), the effect of mortality within the household is both large and statistically significant and the impact of season of birth more than doubles. The results in column 2 imply that a child born into a household in which all children survive will live 2.5 years longer (conditional on dying between age 70 and age 95) than a child born into an otherwise identical household in which only half of the children survive. In the project proposed here, subsequent work will examine both of these effects: for example, does the effect of seasonality vary according to the latitude of the place of birth, which would be consistent with an explanation based on weather and reduced maternal nutrition during pregnancy? And what factors are responsible for the effect of household mortality on longevity?

2. Household Structure in 1920, Physical Growth to Age 25, and Longevity

In order to assess the impact of early-life conditions for the cohorts born just before 1920 on longevity and on characteristics around age 25 (at enlistment in the military), males born 1915-1919 in the SSDI and males born 1915-1919 in the U.S. Army Serial Number File (which includes measurement of height and weight) were sought in the manuscript schedules of the 1920 U.S. Census of Population. Of these, 24,993 were linked from the SSDI to the census that records their circumstances when they were under age 5 and 5,836 were linked from the Army records to the same census. Table 6 presents OLS regression results on age at death conditional on death between the ages of 57 and 83 and on physical characteristics at enlistment. Not shown are controls for year of birth, month of birth, and state of residence in 1920.

The most striking findings are the effects of absent parents. A missing father when the individual was under age 5 was associated with a life span shorter by nearly a half year, while a missing mother exacted no longevity penalty. In height at enlistment, however, an absent mother was associated with a shortfall in height of more than a half inch, while a missing father had no discernable effect on height. A missing father did have an impact on weight at enlistment (nearly 5 pounds), while an absent mother had no effect. It is tempting to interpret these findings (mothers effect height, fathers effect weight) as evidence that mothers have a greater impact on early-life nutrition than fathers but that an absent father leads to lower nutrition, perhaps because of greater labor force participation requirements in households lacking a male breadwinner (if adult height is taken as a marker for early net nutrition and adult weight is taken as a measure of current net nutrition). Further examination of these pathways will be possible through examination of the household structure to determine when the mother's absence occurred (by examining the ages of children in the household) and by examining the employment of individuals at the time of enlistment.

	(1) Age at Death (days)	(2) Height at Enlistment (inches)	(3) Weight at Enlistment (pounds)	(4) 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)
Non-Migrant	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 City	-18.655	-0.071	1.606**	0.279***
	(38.987)	(0.081)	(0.666)	(0.089)
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 6. OLS Regressions on Age at Death, Males Linked from 1920 Census to SSDI or Army Enlistment Records Who Died Age 57-83.

3. The Impact of Lead

Water systems throughout the U.S. used significant quantities of lead in their water pipes at the start of the twentieth century (Troesken, 2006). We now know a great deal about the long-term consequences of lead exposure: changes in brain structure (Cecil *et al.*, 2008) that can have an impact of behavior, including criminal activity and can lead to substantial changes in IQ (Reyes, 2007). There have been few source with which these impacts could be assessed until now. In particular, there are few large-scale sources with both data on measured intelligence and data on the physical environment within which childhood was spent. The linkage of death records like the SSDI or state death certificates and census sources for the early twentieth century, together with existing sources that describe the characteristics of local water systems in place at the same time, satisfy the second of these needs. Satisfying the first is more difficult, as there are few sources of IQ test data available for large populations before the 1950s, and even fewer with sufficient identifying information to allow linkage to sources such as the census and death records.

We have already begun the analysis of the information on height and weight in the U.S. Army's Serial Number file for enlistments during World War Two (see section 3 above). We were aware that, as part of the process of assigning individuals to specific service branches and tasks in the

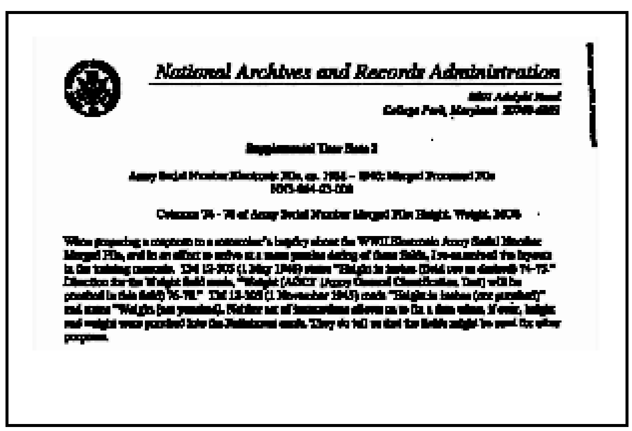
National Archives and Records Administration

FIGURE 6. National Archives Documentation for the U.S. Army Serial Number Electronic File.

military, the Army made use of an early IQ test, the Army General Classification Test (AGCT) that was administered at the time of enlistment. Feyrer *et al.* have exploited the fact that individuals assigned to the Army Air Corp (the forerunner of the modern U.S. Air Force) were required for much of the war to be among the highest scorers on the AGCT, together with the reported branch assignments in the enlistment file, to create a proxy for IQ at enlistment: individuals assigned to the Army Air Corp are assumed to have scored higher on the AGCT than other enlistees. But they were unable to take advantage of the individual-level test scores themselves. In fact, no researchers until now were aware that the scores were actually reported in the enlistment file. Figure 6 shows a copy of the National Archives documentation for the file, based on the field instructions (training manual TM 12-305, May 1, 1943) to clerks at enlistment sites responsible for keypunching the enlistment data. The instructions for the "weight" field read: "Weight (AGCT will be punched in this field) 76-78." This suggests that for at least some time around the start of May, 1943, the enlistment file contains the AGCT results in columns 76-78.

In order to confirm this conjecture, we examined the information actually present in the "weight" field by week for each of the enlistment sites in the file. In most cases, at the beginning of March, 1943, the distribution of values in this field shifts noticeably: before March, 1943, it is an essentially normal distribution centered at 147 with a standard deviation of 20, but in March it shifts to a left-skewed distribution with a mean of 97 and a standard deviation of 24. The pre-March distribution corresponds to the known distribution of weight among enlistees (Karpinos, 1958); the distribution for March, April, May, and the first weeks of June corresponds instead to the know distribution of AGCT scores among enlistees (Staff, Personnel Research Section, the Adjutant General's Office, 1947). Figure 7 shows the distributions for two months: November, 1942 and May, 1943.

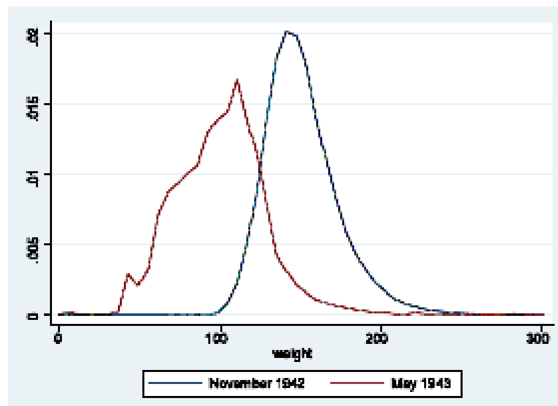

FIGURE 7. Distributions of Values in the "Weight" Field in the Army Serial Number Electronic File, 1938-46.

We have taken this as conclusive evidence that, at least for 13-14 weeks in the first half of 1943, AGCT scores are available on an individual basis in the enlistment file. This period includes the enlistment of 523,956 individuals. Together with the SSDI and the 1920 census, this makes it possible to assess both (1) the link between early conditions (such as exposure to water-borne lead) and IQ at roughly age 25, and

	Baseline (1)	No Lead (2)	Lead (3)	Lead & Mover (4)	Lead & Stayer (5)
Intercept	99.15	101.99	98.16	101.17	95.03
	(6.28)***	(13.70)***	(7.18)***	(10.82)***	(9.94)***
Water Hardness	0.01	0.00	0.01	0.01	0.02
	(0.01)**	(0.01)	(0.01)**	(0.01)	(0.01)**
Water Acidity	-1.37	-1.90	-1.19	-1.53	-1.25
	(0.82)*	(1.91)	(0.91)	(1.40)	(1.21)
Adjusted R^2	0.0922	0.0757	0.0974	0.1099	0.0940
Observations	2,009	562	1,447	747	700

Standard errors in parentheses. p-values * < 10% ** < 5% *** < 1%

TABLE 7. OLS Regressions on AGCT Score (IQ) at Enlistment, Males Born 1915-19 Located in the 1920 Census in Cities and Town With Water Supply Information.

(2) the link between IQ at roughly age 25 and longevity. The results of preliminary analyses along these lines are shown in Table 7 and 8.

As Troesken (2006, pp.23-140) reports, the amount of lead in water provided by a system that employs lead pipes will vary with the chemical composition of the water itself: more lead will be present in a system that transports soft, acidic water than in one that transports hard, basic water. The analysis in Table 7 is limited to individuals born 1915-19 with reported AGCT scores who were residents in 1920 (when they were under age 5) of places for which we have been able to ascertain the type of service mains used in the water system. Not shown are controls for race, household size, missing parents, birth order, year and quarter of birth, and city size. Column 1 reports the impact of water composition for all such individuals. In column 2 and 3, the sample is limited to individuals in places without and with lead service mains, while in columns 4 and 5, the sample from places with lead service mains is further divided by whether the individual's residence at enlistment was different from ("Mover") or the same as ("Stayer") the 1920 location.

The results, even with these small samples, reveal a negative impact from lead exposure that is largely confined to places that had not just lead pipes but also soft, acidic water. Though the coefficient on water acidity does not achieve statistical significance, its sign is negative as the theory predicts; the coefficient on water hardness is statistically significant in columns 3 and 5 but not in columns 2 and 4, indicating that not living in a place with lead pipes at all yields no negative impact from water softness, while even living in a place with lead pipes but moving by age 25 can mitigate the impact of water softness and lead pipes on IQ. Finally, the magnitude of these effects can be compared to the impact of modern efforts to eradicate atmospheric lead exposure. Reyes (2007) reports that the scientific consensus is that with a fall in blood lead levels of 15 μg/dL (as actually occurred 1976-1990 when atmospheric lead exposure was reduced), IQ scores in the U.S. rose 7.5 points. Moving from the 5th to the 95th percentiles in hardness and from the 95th to the 5th percentiles in acidity (i.e. to harder, less acidic water) and using the coefficients in column 5 leads to an increase in IQ of 6.0 points, nearly as great as the gain in IQ from the elimination of leaded gasoline and leaded paint.

Though we do not yet possess the income and disability data from the NUMIDENT file with which to assess the impact of IQ on subsequent labor market performance, we can assess the association between IQ and longevity for individuals linked from the Army records to state death records. This is easier than linkage to the SSDI, as state death records often report state of birth, an additional characteristic that can be used for record linkage. The California death records are particularly useful, as they span the interval 1945 to 2004. As a result, the "window" within which deaths must occur for them to be observed is wider than in other states (this information is exploited further, along with the ability to follow women as well, in section 5). The analysis reported in Table 8 is limited to individuals who had

	(1)	(2)
Intercept	22722.00	22854.00
	(183.26)***	(189.43)***
AGCT Score	8.67	6.74
	(1.42)***	(1.61)***
High School Graduate		104.45
		(78.38)
Some College		193.03
		(114.71)*
College Graduate		433.25
		(172.65)**
Adjusted R^2	0.0034	0.0047
Observations	13,751	13,751

Standard errors in parentheses. p-values * < 10% ** < 5% *** < 1%

TABLE 8. OLS Regressions on Age at Death (in Days), Males Born 1915-1926 Who Died in California Between Ages 35 & 78.

reported AGCT scores and who died in California between the ages of 35 and 78. Not shown are controls for race, year and quarter of birth, and migration either from birth to enlistment or from enlistment to death.

The results in column 1 reveal a substantial effect of IQ on longevity: a one standard deviation increase in IQ is associated with an addition of nearly half a year to an individual's life span. The Army enlistment file also reports the individual's educational attainment at the time of enlistment (some state death records also report educational attainment at the time of death, which we will exploit at a later date), making it possible to see whether IQ has an effect independent of the greater educational attainment it may permit. The results in column 2 show that even when educational attainment is controlled for (the omitted category is "Did Not Graduate High School"), IQ measured by the AGCT is associated with an effect on longevity that remains substantively large and statistically significant: on additional standard deviation on the AGCT is associated with an addition third of a year in longevity.

A small number of studies (Deary and Batty, 2006) have examined the direct link between IQ and longevity. Whalley and Deary (2001) found a positive link in a sample of 2,230 individuals born in 1921 who attended school in Aberdeen, Scotland where they were administered an IQ test in 1932 and whose vital status was known in January, 1997. The authors of that study suggest four possible mechanisms as sources of this link: (1) childhood IQ predicts later life healthy behaviors; (2) childhood IQ predicts the ability to obtain a less dangerous occupation; (3) childhood IQ acts a record of the bodily "insults" suffered by the individual up to the date of the examination; and (4) childhood IQ acts as "as a general, moderately stable, indicator of system integrity within the body by indexing the efficiency of information processing in the nervous system." Two of these pathways can be investigated further with the data the present project will generate: we will know (from the enlistment file and state death records) the occupations pursued at age 25 and later in life (hypothesis 2); and we will know later life disability and the exact cause of death and whether it is consistent with degeneration of the nervous system (hypothesis 4). Hemmingsson *et al.* (2006) examined mortality over a 30 year follow-up of Swedish males whose IQ was tested at induction into military service (ages 18-20) and found a strong association between IQ and both all-cause mortality and specific causes of death (cardiovascular disease, violence, and alcohol-related deaths), effects that were not eliminated when controls for adult socio-economic status were introduced. No studies to date have examined the relationship between IQ and longevity for the U.S.

4. Differences By Gender and Age in the Impact of Early Conditions on Later Outcomes

The analyses conducted to this point have two limitations: (1) they focus exclusively on males (because the initial linkages we have done to late-life sources such as the SSDI can be done most readily for males, until we receive the full set of identifying information including women's maiden names from SSA's NUMIDENT file); and (2) they focus on deaths that occur within a relatively narrow "window" (a necessary restriction given that the SSDI runs only from 1965 to 2006). In order to assess how the impact of early conditions on longevity differs by gender and by age at death, we have taken advantage of the availability of both birth and death records for California. Because the death records contain exact date of birth and allow us to identify individuals who were born in California and died in California, we are able to link a substantial number of birth and death records for both women and men by conducting the linkage without regard for the reported surname – simply using given name, the exact date of birth (month, day, and year), and the fact that both the birth and the death occurred in California. Furthermore, because the California death records with this information extend back to 1945, it is possible to widen the "window" within which deaths are observed. These individuals linked from California death records back to California birth records (for 1915-19) were then located in the 1920 U.S. Census of Population by searching for them exclusively within California (though this misses individuals who were born in California, moved out of the state between birth and 1920, and then moved back into the state before death), in order to examine their early-life household conditions.

Table 9 presents the results of OLS regressions on longevity for 6 groups: (1) men who died between ages 25 and 85; (2) men who died between ages 25 and 60; (3) men who died between ages 61 and 85; (4) women who died between ages 25 and 85; (5) women who died between ages 25 and 60; (6) women who died between ages 61 and 85. Several differences by gender and age at death are apparent. For women who died between 25 and 85, season of birth does not influence longevity; for men who died between 25 and 85, however, season of birth has a pronounced effect, but not the "second quarter penalty" seen elsewhere. Instead, men born in the fourth quarter had the shortest lives. When the male sample is divided by age at death, the largest effects for quarter of birth come among those who died at younger ages (25-60), though the same general pattern (longer lives among those born in the second and third quarters) persists even among those dying at older ages (61-85). Among women, the standard finding of short lives among second quarter births again emerges when deaths at older and younger ages are examined separately: a large penalty (one and a half years) is paid for birth in the second quarter among deaths between 26 and 60. These differences in the season of birth effect by gender and age suggest that there may be fundamentally different mechanisms at work in these subpopulations. The effect of year of birth also differs across gender and age-at-death groups: for men, it appears only for younger deaths, and favors 1915 births (the excluded category in the regression) by substantial margins; for women, the effect is apparent among deaths between 25 and 85, and the effect is reversed, with 1915 the worst year in which to be born, even though the cohorts examined here include those *in utero* at the time of the 1918-19 influenza pandemic. The only effect of an absent parent is for males who died between 25 and 60 and whose father was not present in 1920; the effect for women who died in this age range is not statistically significant. Higher order births were associated with greater longevity among males who died at younger ages but not among females, and urban residence in 1920 had opposite effects for men and women: it increased longevity for men dying between 25 and 60, but reduced it for women dying between 25 and 85.

Taken together, these results suggest the potential usefulness of analyzing the impact of early conditions on later outcomes separately for men and women, and where possible to disaggregating as well by age at death. A puzzle left to be explained is the mechanism underlying these sharp differences.

	Males			Females		
	Died at 25-85	Died at 25-60	Died at 61-85	Died at 25-85	Died at 25-60	Died at 61-85
	(1)	(2)	(3)	(4)	(5)	(6)
Constant	24,418	17,110	27,076	25,152	18,305	27,299
	(167.28)***	(233.81)***	(96.83)***	(220.67)***	(360.05)***	(129.13)***
Born Jan.-Feb.	269.79	147.58	75.81	-17.17	-270.79	99.10
	(115.51)**	(161.44)	(66.87)	(150.83)	(243.33)	(88.47)
Born Apr.-June	291.94	252.97	132.42	-46.29	-492.55	105.54
	(113.39)***	(157.46)*	(65.77)**	(152.37)	(246.13)**	(89.33)
Born July-Sept.	314.84	356.70	153.43	-48.35	-448.55	106.09
	(112.46)***	(155.80)**	(65.29)**	(147.98)	(238.06)*	(86.81)
Born in 1916	-172.05	-369.22	22.96	345.96	17.56	130.70
	(127.96)	(176.29)**	(74.45)	(169.98)**	(267.79)	(100.38)
Born in 1917	4.70	-412.27	95.19	463.57	29.65	203.17
	(127.99)	(178.31)**	(74.17)	(169.54)***	(268.87)	(99.96)**
Born in 1918	117.07	-291.49	54.88	431.59	514.74	115.97
	(128.51)	(181.28)*	(74.14)	(167.65)***	(263.08)**	(99.11)
Born in 1919	-72.48	-520.88	-47.79	469.53	-332.85	229.22
	(128.66)	(179.86)***	(74.46)	(170.04)***	(271.83)	(100.09)**
Household Size	40.89	-34.63	24.76	39.12	-98.93	-3.47
	(36.46)	(51.98)	(20.98)	(45.67)	(77.52)	(26.46)
Father Missing	-320.10	-515.53	5.97	-74.05	-302.15	-74.14
	(189.34)*	(249.98)**	(112.06)	(264.49)	(440.22)	(153.93)
Mother Missing	-92.91	100.93	-108.16	333.32	439.17	58.52
	(231.24)	(307.47)	(136.54)	(313.71)	(535.47)	(181.53)
Birth Order	-28.17	151.24	-37.96	-28.97	93.29	-14.21
	(43.88)	(62.30)**	(25.28)	(55.44)	(92.71)	(32.22)
Urban	-16.01	381.47	-55.69	-327.52	-109.81	-75.88
	(81.29)	(113.44)***	(47.09)	(107.07)***	(171.87)	(62.93)
Observations	15,328	3,599	11,729	7,809	1,492	6,324
Adjusted R^2	0.001	0.001	0.001	0.001	0.004	0.002

Standard errors in parentheses. p-values * < 10% ** < 5% *** < 1%

TABLE 9. OLS Regressions on Age at Death (in Days), Males and Females Born 1915-1919 Who Died in California Between Ages 25 & 85.

E. FUTURE DIRECTIONS

1. Assess the Biases in the Data and Develop Weights

The linked data will not be representative of the entire population of individuals whose early-life circumstances can be observed in census documents, because of the restrictions on the linkage process imposed by the availability of the Social Security Death Master file, because of the availability of original SS-5 information in SSA's NUMIDENT file, and because of information that is either missing or not correctly transcribed in one of the sources. As a result, it will be necessary to generate a weighting scheme to make possible generalizations from the linked population to the general population. The availability of detailed public use samples from the 1900-30 population census as well as detailed information on decedents in the Mortality Detail Files maintained by the National Center for Health Statistics (1968-present) mean that the disparity between the characteristics of the linked records and the total population can be assessed either early or late in life. Several procedures are available to take a set of marginal frequencies in one dataset and calculate weights for each observation that will allow the weighted data to duplicate the same set of marginal frequencies in another (e.g. Izrael *et al.*, 2000).

2. Analyze the Link Between Early-Life Conditions and Outcomes at Age 25

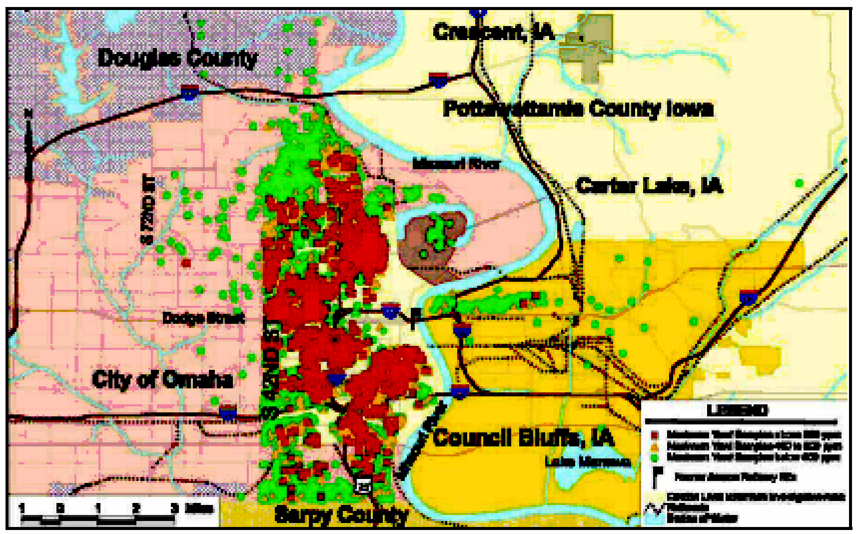

FIGURE 8. Lead Contamination in Omaha, Nebraska, 2004.

Preliminary analysis of individuals linked from U.S. Army enlistment records to the 1920 census of population revealed a strong link between some early-life environmental circumstances and conditions around age 25 (such as residence in a location served by lead water pipes which reduced measured IQ at enlistment if the location has soft, acidic water). The same analysis also identified a birth-order effect on intelligence. Both early-life effects will be explored with a considerably larger body of data. In addition to lead water pipes, other sources of environmental lead contamination will be explored. For example, a large lead smelter operated on the eastern edge of the city of Omaha for 120 years. The Environmental Protection Administration has designated this a priority clean-up site, and has prepared detailed maps that describe the areas in eastern Omaha that received the greatest contamination. Figure 8 shows the locations where soil samples indicate that unsafe quantities of lead are present today. These locations are consistent with prevailing wind patterns and the location of the smelter. By drawing large samples from the 1920 and 1930 population censuses for Omaha, and linking individuals to the enlistment file, it will be possible to evaluate the impact of proximity to the sites that received the greatest contamination on subsequent measured IQ (as well as height, weight, occupation, and educational attainment). A number of other lead smelters have been identified for which similar analyses can be performed.

There are several other early-life conditions that will be investigated as contributing to later IQ: family size, birth order, household SES, school attendance, and the disease environment faced early in life. In each case, the analysis will consist of ordinary least squares or ordered probit regressions in which an outcome at age 25 (IQ, height, weight, education) is examined as a result of a range of early conditions at the individual (season and year of birth, school attendance), household (SES, birth order, family size), and neighborhood (proximity to hazards, local disease environment) levels. The size of the dataset will allow analysis of a variety of interactions among individual, household, and neighborhood effects.

3. Analyze Links Between Early-life Conditions and Later-life Outcomes

A large body of research now implicates early-life conditions in the development of later-life chronic conditions and early mortality (see section "B. Background and Significance" above). Few of these links have ever been explored for large populations in the U.S. or for populations at older ages. The linked data this project will generate will provide several opportunities for such analysis. Two approaches will be pursued: (1) simple OLS regression (for continuous outcomes such as income), probit (for dichotomous outcomes such as disability), or hazard model (for time-dependent processes such as mortality) analyses; and (2) difference-in-differences or regression discontinuity analyses where a discrete "treatment" can be identified and its differential impact either over time or across space can be exploited.

An example of the first approach is examining the role played by the disease environment encountered early in life on the later-life onset of specific health conditions (using the information on specific diagnosis in the Medicare and Veterans Department files and on cause of death in state death records) and the role of household SES in childhood in mediating that role. Another example is the analysis of the link between IQ and later health and longevity which has been identified in several small, non-U.S. populations but never explored with a large body of data, rich household and neighborhood context early in life, and the ability to examine a variety of later-life health indicators. Examples of the second approach that exploits shocks to local or national populations include: assessing the impact on income, disability, health, and longevity of the death of a parent (by examining particular communities like those with large numbers of miners where such events were common enough to generate a large number of "treatments"); assessing the impact of large-scale vaccination and child health examination programs undertaken in some U.S. cities in the 1920s; assessing the impact of the reversal of the flow of the Chicago River (completed in 1900 with the opening of the Sanitary and Ship Canal) on long-term health outcomes for residents of Chicago and St. Louis; and the impact on later health of disruptions to local food supplies and employment resulting from severe weather conditions (tornadoes and hurricanes) in the early twentieth century (we have obtained information on the track of these storms and the communities damaged).

4. Assess Differences Across Groups in the Link Between Early Conditions and Later Outcomes

A significant shortcoming of previous work on the impact of early conditions on later outcomes is the inability to identify these effects for small populations (e.g. particular ethnic or racial groups; communities in different climates) or to disaggregate by sex in the analysis. The data generated in this project will be sufficiently broad that even small sub-populations can be analyzed separately. The preliminary analysis of differences between men and women in the impact of conditions early in life on longevity suggests the usefulness of this approach. Analysis of a variety of outcomes will be conducted in which separate regressions are estimated for distinct population groups and the regression coefficients will be compared.

5. Explore the Role of Parents' (Particularly Mothers') Backgrounds

Gluckman *et al.* (2008, p. 68-69) suggest that the search for the sources of later-life health conditions must be extended back past an individual's own childhood and *in utero* development to the early-life development of the parents, particularly the mother. For individuals observed in the 1920 and 1930 censuses and subsequently linked to later-life outcomes, this is feasible if the individual is also linked to the NUMIDENT file or to a state death record that reveals the maiden name of the individual's mother. In such cases, together with the information on the mother's age and birthplace contained in the census record for 1920 or 1930, it will be possible to link the mother back to the 1900 household in which she herself lived as a child. For example, a mother who was 35 years old in 1930 and had a 5 year old child at that date would herself be observed as a 5 year old in the 1900 census. This would provide the opportunity to examine the link between the mother's own childhood SES and her child's subsequent height, weight, IQ, health, income, and longevity, while controlling for the SES experienced by the child in early life. In addition, because of the presence of the unique questions in the 1900 census on children ever born and children surviving asked of mothers, it would be possible to examine the impact of growing up in a high-mortality household on outcomes for children of someone raised in such a household. The ability to establish these intergenerational links is extremely limited in most modern data sources.

Other aspects of the background of parents and of their interactions can be investigated as sources of differences in outcomes for children later in life. For example, the gap between the ages of husbands and wives stood at its historic high (more than 4 years on average) in 1900 after rising since the 1850s and before falling through the twentieth century to its current level of just over 2 years. These age gaps are thought by some to reflect differences in the relative bargaining positions of husbands and wives as they enter into marriage, differences that can be manifested in intra-household resource allocations. Whether these differences are reflected in much later outcomes for children raised in households where the gap between the father's age and the mother's age was great has not yet been examined. Even if intra-household bargaining is not impacted by these age gaps, parental age gaps may nonetheless have an impact on outcomes: households with older fathers may be more prone to have children with chronic health conditions. We do not yet know how likely these households are to have children with conditions that appear only later in life, or that have an impact on longevity. A countervailing, positive impact from experiencing early life in a household where the father is considerably older than the mother is the combination of better maternal health at younger ages and greater total resources available to the household because the father is closer to the peak of his age-earnings profile (which will be important where credit markets are imperfect and it is difficult for younger fathers to borrow against future labor earnings). These pathways can be investigated with data on parents' ages, household SES, asset ownership, and investments in children early in life (e.g. school attendance).

F. CONCLUSIONS

Though interest in the impact of early-life circumstances on later-life outcomes has increased dramatically in recent years, few available datasets simultaneously permit the examination of these links (1) on a large scale, (2) for a wide range of outcomes, and (3) with detailed neighborhood and family context. The present project explores the usefulness of linking U.S. Census records (1900-30) to Social Security records, state death records, U.S. Army enlistment records, and other sources to produce data that satisfies all of these criteria.

Preliminary analysis reveals that early-life circumstances as mundane as the individual's season of birth or the household or community disease environment have a substantial impact on longevity. Other circumstances faced by individuals before they reach age 5, such as the absence of a parent, have their effect on the individual's height and weight in early adulthood. Exposure to lead in drinking water has marked cognitive effects in early adulthood that are comparable in size to the impact of lead exposure from gasoline and paint, and cognitive functioning in early adulthood, in turn, has a substantial impact on longevity. Further investigation of these effects will be undertaken to assess the specific pathways through which they operate. In addition, other outcomes (income and health at older ages, specific cause of death) will be examined to assess how they are linked to these early-life conditions.

G. BIBLIOGRAPHY

Almond, Douglas (2006). "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*, 114:4 (August), 672-712.

Barker, D.J.P. (ed.) (1992). *Fetal and Infant Origins of Adult Disease: Papers* (London: British Medical Journal).

Barker, D.J.P. (1998). *Mothers, Babies, and Health in Later Life* (Edinburgh: Churchill Livingstone).

Barker, D.J.P., K.M. Godfrey, C. Fall, C. Osmond, P.D. Winter, and S.O. Shaheen (1991). "Relation of Birth Weight and Childhood Respiratory Infection to Adult Lung Function and Death from Chronic Obstructive Airways Disease." *British Medical Journal*, 303, 671-675.

Barker, D.J.P., C. Osmond, S.J. Simmonds, and G.A. Wield (1993). "The Relation of Small Head Circumference and Thinness at Birth to Death from Cardiovascular Disease in Adult Life." *British Medical Journal*, 306, 422-426.

Becker, Gary S., and Nigel Tomes (1976). "Child Endowments and the Quantity and Quality of Children." *Journal of Political Economy*, 84:4 (August), part 2, S143-S162.

Black, Sandra E., Paul J. Devereux, and Kjell G. Salvanes. "From the Cradle to the Labor Market? The Effect of Birth Weight on Adult Outcomes." *Quarterly Journal of Economics*, 122:1 (February), 409-439.

Case, Anne, Angela Fertig, and Christine Paxson (2005). "The Lasting Impact of Childhood Health and Circumstance," *Journal of Health Economics*, 24: 2 (March), 365-389.

Cecil, Kim M., Christopher J. Brubaker, Caleb M. Adler, Kim N. Dietrich, Mekibib Altaye, John C. Egelhoff, Stephanie Wessel, Ilayaraja Elangovan, Richard Hornung, Kelly Jarvis, and Bruce P. Lanphear (2008). "Decreased Brain Volume in Adults with Childhood Lead Exposure." *PLoS Medicine*, 5:5 (May), 741-749.

Costa, Dora L., and Joanna N. Lahey (2005). "Predicting Older Age Mortality Trends." *Journal of the European Economic Association*, 3 (April–May) 487–93.

Currie, Janet, and B.C. Madrian (1999). "Health, Health Insurance and the Labor Market." In O. AshenJOlter and D. Card (eds.), *Handbook of Labor Economics*, 3:3 (Elsevier), 3309-3416.

Currie, Janet (2008). "Healthy, Wealthy, and Wise: Socioeconomic Status, Poor Health in Childhood, and Human Capital Development." *Journal of Economic Literature* (forthcoming).

Deary, I.J., and and G. David Batty (2006). "Commentary: Pre-morbid IQ and later health—the rapidly evolving field of cognitive epidemiology." *International Journal of Epidemiology,* 35:3, 670-672.

Doblhammer, Gabriele and James W. Vaupel (2001). "Lifespan Depends on Month of Birth." *Proceedings of the National Academy of Sciences*, 98, 2934–2939.

Edwards, Thomas I. , and Louis P. Hellman (1944). "Methods Used in Processing Data from the Physical Examination Reports of the Selective Service System." *Journal of the American Statistical Association*, 39: 226 (June), 165-182

Elo, Irma T. . Cassio M. Turra, Bert Kestenbaum, B. Renee Ferguson (2004). "Mortality among Elderly Hispanics in the United States: Past Evidence and New Results." *Demography*, 41:1 (February), 109-128.

Eriksson, J.G., T. Forsen, J. Tuomilhto, P.D. Winter, C. Osmond, and D.J.P. Barker (1999). "Catch-Up Growth in Childhood and Death from Coronary Heart Disease: Longitudinal Study." *British Medical Journal*, 318, 427-431.

Eriksson, J.G., T. Forsen, J. Tuomilhto, C. Osmond, and D.J.P. Barker (2001). "early Growth and Coronary Heart Disease in Later Life: Longitudinal Study." *British Medical Journal*, 322, 949-953.

Eriksson, J.G., T.J. Forsen, C. Osmond, and D.J.P. Barker (2003). "Pathways of Infant and Childhood Growth That Lead to Type 2 Diabetes." *Diabetes Care*, 26, 3006-3010.

Feyrer, James, Dimitra Politi, and David N. Weil (2007). "The Economic Effects of Micronutrient Deficiency: Evidence from Salt Iodization in the United States." Working Paper (December).

Forsen, T., J.G. Eriksson, J. Tuomilehto, K. Teramo, C. Osmond, and D.J.P. Barker (1997). "Mother's Weight in Pregnancy and Coronary Heart disases in a Cohort of finnish Men: Follow-Up Study." *British Medical Journal*, 315, 837-840.

Frankel, S., P. Elwood, P. Sweetnam, J. Yarnell, and G. Davey Smith (1996). "Birthweight, Body-Mass Index in Middle Age, and Incident Coronary Heart Disease." *Lancet*, 348, 1478-1480.

Gale, C.R., S. Walton, and C.N. Martyn (2003). "Foetal and Postnatal Head Growth and Risk of Cognitive Decline in Old Age." *Brain*, 126, 2273-2278.

Gale, C.R., F.J. O'Callaghan, K.M. Godfrey, C.M. Law, and C.N. Martyn (2004). "Critical Periods of Brain Growth and Cognitive Function in Children." *Brain*, 127, 321-329.

Gluckman, Peter D., and Mark A. Hanson (2006). "The Developmental Origins of Health and Disease: An Overview." In Peter D. Gluckman and Mark A. Hanson (eds.), *Developmental Origins of Health and Disease* (Cambridge: Cambridge University Press).

Gluckman, Peter D., Mark A. Hanson, Cyrus Cooper, and Kent L. Thornburg (2008). "Effect of In Utero and Early-Life Conditions on Adult Health and Disease." *New England Journal of Medicine*, 359:1 (July 3), 61-73.

Godfrey, Keith (2006). "The Developmental Origins of Health and Disease: Epidemiology." In Peter D. Gluckman and Mark A. Hanson (eds.), *Developmental Origins of Health and Disease* (Cambridge: Cambridge University Press).

Hayward, Mark D. , and Bridget K. Gorman (2004). "The Long Arm of Childhood: The Influence of Early-Life Social Conditions on Men's Mortality." *Demography*, 41: 1 (February, 2004), 87-107 .

Harvey, N., and C. Cooper (2004). "The Developmental Origins of Osteoporotic Fracture." *Journal of the British Menopause Society*, 10, 14-29.

Heckman, James J. (2008). "Schools, Skills, and Synapses." National Bureau of Economic Research Working Paper No. 14064 (June).

Hemmingsson, T., B. Meli, P. Allebeck, and I.Lundberg (2006). "The Association Between Cognitive Ability Measured at Ages 18–20 and Mortality During 30 Years of Follow-Up: a Prospective Observational Study among Swedish Males Born 1949–1951." *International Journal of Epidemiology*, 35, 665–670.

Huxley, R.R., A.W. Shiell, and C.M. Law (2000). "The Role of size at Birth and Post-Natal Catch-Up Growth in Determining Systolic Blood Pressure: A Systematic Review of the Literature." *Journal of Hypertension*, 18, 815-831.

Izrael, David, Hoaglin, David C., and Battaglia, Michael P. (2000), "A SAS Macro for Balancing a Weighted Sample." *Proceedings of the Twenty-Fifth Annual SAS Users Group International Conference*, Paper 275.

Johnson, Rucker C., and Robert F. Schoeni (2007). "The Influence of Early-Life Events on Human Capital, Health Status, and Labor Market Outcomes Over the Life Course." Institute for Research on Labor and Employment, Population Studies Center Research Report 07-616 (January).

Karpinos, Bernard D. (1958). "Height and Weight of Selective Service Registrants Processed for Military Service During World War II." *Human Biology*, 30:4 (December), 292-321.

Lithell, H.O., P.M. McKeigue, L. Berglund, R. Moshen, U.B. Lithell, and D.A. Leon (1996). "Relation of Size at Birth to Non-Insulin Dependent Diabetes and Insulin Concentrations in Men Aged 50-60 Years." *British Medical Journal*, 312, 406-410.

Margolis, Rachel (2008). "The Effects of Early Childhood Diseases on Young Adult Health in Guatemala." PARC Working Paper Series, No.7 (August).

Martyn, C.N., D.J.P. Barler, and C. Osmond (1996). "Mothers' Pelvic Size, Fetal Growth, and Death from Stroke and Coronary Heart Disease in Men in the U.K." *Lancet*, 348, 1264-1268.

Osmond, C., D.J.P. Barker, P.D. Winter, C.H.D. Fall, and S.J. Simmonds (1993). "Early Growth and Death from Cardiovascular Disease in Women." *British Medical Journal*, 307, 1519-1524.

Panis, Constantijn, Roald Euller, Cynthia Grant, Melissa Bradley, Christine E. Peterson, Randall Hirscher, and Paul Steinberg (2000). *SSA Program Data User's Manual*. Project Memorandum Prepared for the Social Security Administration (Rand Corporation).

Preston, Samuel H. ,Mark E. Hill and Greg L. Drevenstedt (1998). "Childhood Conditions That Predict Survival to Advanced Ages among African-Americans." *Social Science Medicine*, 47:9, 1231-1246.

Reyes, Jessica Wolpaw (2007). "Environmental Policy as Social Policy? The Impact of Childhood Lead Exposure on Crime." *The B.E. Journal of Economic Analysis & Policy*, 7:1 (Contributions).

Rich-Edwards, J.W., M.J. Stampfer, J.E. Manson, *et al.* (1997). "Birth Weight and the Risk of Cardiovascular Disease in a Cohort of women Followed Up Since 1976." *British Medical Journal*, 315, 396-400.

Ruggles, Steven, Matthew Sobek, Trent Alexander, Catherine A. Fitch, Ronald Goeken, Patricia Kelly Hall, Miriam King, and Chad Ronnander (2008). *Integrated Public Use Microdata Series: Version 4.0* [Machine-readable database] (Minneapolis, MN: Minnesota Population Center [producer and distributor]).

Sacker, A., D.J. Done, T.J. Crow, and J. Golding (1995). "Antecedents of Schizophrenia and Affective illness: Obstetric Complications." *British Journal of Psychiatry*, 166, 734-741.

Singhal, A., I.S. Farooqi, S. O'Rahilly, T.J. Cole, M. Fewtrell, and A. Lucas (2002). "Early Nutrition and Leptin Concentrations in Later Life." *American Journal of Clinical Nutrition*, 75, 993-999.

Smith, George Davey, and Shah Ebrahim (2001). "Epidemiology – Is It Time to Call It a Day?" *International Journal of Epidemiology*, 30, 1-11.

Staff, Personnel Research Section, the Adjutant General's Office (1947). "The Army General Classification Test, with special reference to the construction and standardization of Forms 1a and 1b." *Journal of Educational Psychology*, 38, 385–420.

Stein, Z. A., M. Susser, G. Saenger, and F. Moraolla (1975). *Famine and Human development: The Dutch Hunger Winter of 1944–45* (New York, Oxford University Press).

Susser, E., R. Neugebauer, H.W. Hoek, *et al.* (1996). "Schizophrenia After Prenatal Famine:

Further Evidence." *Archives of General Psychiatry*, 53, 25-31.

Troesken, Werner (2006). *The Great Lead Water Pipe Disaster* (Cambridge: MIT Press).

Trzcinka-Ochocka, M., M. Jakubowski, G. Razniewska, T. Halatek, and A. Gazewski (2004). "The Effects of Environmental Cadmium Exposure on Kidney Function: the Possible Influence of Age." *Environmental Research*, 95:2 (June), 143-150.

Villermé, L.R. (1829). "Memoire sur la taille de l'homme en France." *Annales d'Hygiene Publique*, 1, 551-559.

Whalley, Lawrence J., and Ian J. Deary (2001). "Longitudinal Cohort Study of Childhood iq and Survival up to Age 76." *British Medical Journal*, 322: 7290 (April 7).