Race, Socioeconomic Status, and Mortality in the 20th Century: Evidence from the Carolinas

Trevon D. Logan*
Department of Economics,
The Ohio State University and NBER
410 Arps Hall
1945 N. High Street
Columbus, OH 43210
logan.155@osu.edu

John M. Parman
Department of Economics,
University of California, Davis
One Shields Avenue
Davis, CA 95616
jmparman@ucdavis.edu

August 5, 2010

Abstract

Racial and socioeconomic gaps in mortality persisted throughout the twentieth century. We know little, however, about racial or socioeconomic gaps in cause-specific mortality or in how the two are related over time. Demographers have repeatedly documented serious data problems that limit our ability to analyze these issues. In an attempt to overcome these problems, we link a random sample of death certificates taken at five year intervals from 1910 to 1975 to the manuscript federal census files of the deceased early in life and then to the death certificates of the deceased's parents. To our knowledge, the data we construct is the first of its kind by linkng parent and child death certificate information with the additional information from the census files. We show that our research design allows us to construct a panel data set that allows us to look at mortality (both general and cause specific) over time and for specific cohorts. This paper presents preliminary evidence from our pilot study of death certificates from the Carolinas in the twentieth century. We also outline several avenues of future research to be investigated with this data.

JEL classifications: I1, J1, N3

* We thank George Alter, Susan Leonard, and Bernardo Queiroz for extremely helpful conversations. We acknowledge the financial support of the Robert Wood Johnson Foundation and a Arthur H. Cole Grant from the Economic History Association. The usual disclaimer applies.

1 Introduction

Racial and socioeconomic differences in mortality persisted throughout the twentieth century and have continued into the twenty first. The progress made in closing the gap has been uneven– while the mortality rate of black men did not fall below that of white men in 1900 until 1975, gaps in black-white infant mortality fell further over a shorter period of time [Costa 2005, Bell and Miller 2005, Ewbank 1987, Levine et al 2001, Zelnik 1969, National Office of Vital Statistics 1956]. Recent demographic scholarship continues to document trends in racial mortality differences [Harper et al 2007, Elo and Drevenstedt 2004, Geronimus et al 2001, Lynch 2008, Brown and Lynch 2004]. We know little, however, about the mortality trends before 1970, especially for features other than the general mortality rate. A large part of the problem is data– birth and death registrations early in the century are incomplete and official counts of the African American population and number of deaths in that population are known to be biased [Eblen 1974, Coale and Rives 1973, Elo 2001, Elo and Preston 1994, Preston at al 1998, Rosenberg et al 1999]. Researchers have also documented significant measurement error in black ages among the aged, making inference about racial differences in older age mortality, precisely where mortality is concentrated, difficult [Elo et al 1996, Hill et al 1997]. Indeed, the "mortality crossover," where at older ages the mortality of blacks has been shown to be *lower* than whites, has been challenged as a figment of age misreporting among the African American population [Coale and Kisker 1986, Preston et al 1996, Rosenwaike and Hill 1996]. Others, however, argue that the finding is robust and extends to specific causes of death for the late twentieth century [Lynch et al 2003, Eberstein and Nam 2008]. To date, no consensus has been reached, and we know little about the potential for a cross over in the past other than the general mortality rate [Elo 2001].

Unfortunately, demographic research cannot escape the racial stain of the past: while whites are found to have extremely low rates of age misreporting and generally excellent population coverage throughout the twentieth century [Rosenwaike and Logue 1983, Hill et al 2000], our historical demographic data on the African American population is lacking. For example, Elo [2001] notes that there exist no official lifetables for the black population from 1935 to 1970. This inhibits our ability to look beneath the surface of mortality rates themselves. Due to the lack of data we have few studies that look at racial disparities by cause of death in the past [Costa 2005, Costa et al 2007],

and therefore know little about convergence or divergence in black and white health as opposed to mortality. The vast majority of studies that explore the racial gaps in specific health conditions are contemporary. The small number of studies that do investigate cause-specific differences in the past have found significant racial differences. Costa et al [2007] find significant differences in the rates of heart disease for black and white men early in the 20th century, and Costa [2005] attributes much of the difference between black and white older age mortality in the early twentieth century to differences in infectious diseases, which in related work can be seen in birth outcomes and sanitation improvements in the South as well [Troesken 2004, Costa 2004]. A drawback of these studies is that they apply to a highly specialized cohort (Civil War veterans), who may be different from the general population on a host of observable and unobservable measures. Similarly, those born in the antebellum era ended their lives in the early twentieth century, which limits our ability to learn about changes that have taken place after the advent of large scale public health measures and the dramatic social and economic changes of the twentieth century. Another drawback is that possbile mechanisms for the mortality convergence, key for policy analysis, cannot be tested in the existing data. Troesken [2004], for example, argues that improvements in sanitation led to significant declines in black mortality, but the stronger empirical test would be to look for declines in cause-specific mortality that would be directly related to sanitation. Others have adopted regression-discontinuity and differences-in-differences estimation strategies to identify the effects of hospital availability and desegregation in explaining black/white mortality differences, but more specific test of whether conditions that would be more sensitive to access to hospitals is lacking [Almond et al 2006, Almond, Chay and Greenstone 2006]. In general, if infectious disease differentials or acces to care were the cause of a significant portion of the black/white mortality gap in the past then declines in the racial mortality gap due to infectious disease causes of death or causes of death sensitive to hospital access should be larger than for other causes. Unfortunately, empirical tests of these types of hypotheses have not been possible on a large scale.

Race is not the only dimension of mortality differences, however. Since Kitagawa and Hauser's [1973] landmark study of socioeconomic differences in mortality researchers have found mortality gaps in a host of dimensions such as education [Elo and Preston 1996], occupation [Moore and Hayward 1990, Mare 1990] and childhood socioeconomic status [Haywward and Gorman 2004]. Unlike racial gaps in mortality, the socioeconomic gaps have arguably been growing over time

[Lauderdale 2001, Lynch 2003, Pappas et al 1993, Hummer et al 1998]. Researchers have recently begun to look at the interaction of race and socioeconomic status in explaining mortality differences [Rogers 1992]. For example, researchers have found that socioeconomic status explains a significant part of the racial gap in chronic health conditions [Hayward et al 2000] and mortality [Rogers 1992, Sudano and Baker 2005]. Preston et al [1998] find that early life socioeconomic status explains a significant portion of racial differences in mortality. What is not known is the degree to which other factors play a role. For example, given the large differences in mortality by geography in the early part of the last century that varied with macrosocioeconomic measures [Crimmins and Condran 1983], and the unequal distributions of the black and white populations, part of the differences we attribute to "early life circumstances" may be better explained as environmental or social differences rather than socioeconomic per se [see Naidu 2009]. Even more, the relationships underlying these trends are inherently dynamic and can make it difficult to identify specific mechanisms. Su [2009], for example, finds occupational gradients in later life mortality for cohorts born in the 1840s favored those who did not leave their earlier occupations as farmers, which is consistent with Cuff's [2005] and Yoo's [2010] argument that those of lower socioeconomic status enjoyed a health advantage primarily due to geography. Logan [2009] argues that migration and socioeconomic mobility where related to health and slave status in the late nineteenth century for African Americans, implying that part of the racial differences we observe are due to socioeconomic differences that themselves could be due to health differences.

Researchers have hypothesized that socioeconomic disadvantages in health are cumulative [O'Rand 2002], owing to socioeconomically based risk factors and differential access to resources that could promote health. Researchers have also hypothesized that racial differences in health increase later in life due to the "double jeopardy" of age and race [Ferraro 1987, Brown and Lynch 2004]. Others hold that racial and socioeconomic differences decline at older ages as biological factors outweigh the social factors– at older ages we are left with healthy individuals irrespective of their racial of socioeconomic origins. Tests of these competing hypotheses requires information on socioeconomic status and cause of death over time. Little research looks to see how socioeconomic and racial differences may interact with one another over age, cohorts, and time.

Beyond this, we have little empirical evidence on the degree of intergenerational transmission of health and its role in explaining racial and socioeconomic gaps in mortality. The early life

circumstances of children largely reflect the mid-life circumstances of parents, which could be related, naturally, to the early life circumstances of the parents. Recent research has documented that the socioeconomic gradient in health develops early and persists, and the development of chronic conditions in poor households has more serious health consequences than those in wealthy households [Case et al 2002]. Indeed, the contemporary evidence suggest that low income children are more likely to be subject to health shocks [Currie and Stabile 2003]. We know little about the historical dynamics of such relationships. For example, if a significant portion of black-white convergence in mortality was due to better socioeconomic conditions of parents then the effects of public policies such as antidiscrimination laws has been underestimated. Similarly, as the returns to education increased over the twentieth century the intergenerational transmission could be stronger if healthier parents raise healthier children who are better able to sufficiently invest in their education [Goldin and Katz 2008, Andrews and Logan 2010]. On the other hand, the story of the twentieth century in general was one of declining mortality and the arguable conquest of infectious disease in the United States, we may expect the effects of early life circumstances to decline over time [Fogel 2004]. Naturally, this could differ by race for a host of reasons. And similarly, data from the past are generally lacking to estimate such relationships.

To better analyze the trends in racial and socioeconomic differences in mortality over time we would need measures of individuals from a variety of socioeconomic and racial backgrounds and information on their early life circumstances and later diseases and death. Lynch [2008] argues that the ideal data to answer these inherently dynamic questions would be a panel that includes measures of socioeconomic status and covering many birth cohorts. While Lynch describes contemporary panel data that has some of those desirable features, the fact that these processes evolve over time necessitates that we develop methods to investigate these patterns in the past. The approach that we take with this project is one step in that direction.

In this paper we present preliminary results from a pilot project which seeks to add to our ability to describe the trends in mortality in the twentieth century. Building on the pioneering matching methodology described by Preston et al [1996] and the general techniques laid out by Fogel [1993], we obtain a random sample of death certificates from North and South Carolina at five year intervals from 1910 to 1970. Since death certificates also list the place of birth and parental information, we use those pieces to match the deceased to their childhood home in the manuscript

federal census, giving us a rich set of measures of early life circumstances as well as an independent measure of age. Moving beyond the matching to correct for age misreporting, we exploit the rich data in the census which gives us measures of the socioeconomic status of the household in the early life of the now-deceased. We also advance the field by using the detailed information on the death certificates, which list cause of death, allowing us to look at racial and socioeconomic differences in cause specific mortality over the twentieth century, and use the ages in the census to correct for age misreporting on death certificates. Further, we match the parents of the deceased in our data to their own death certificates, giving us both parent and child death certificates. To our knowledge, our linking of death certificates to federal census records and then to parental death certificates is the first of its kind. We use this linking combined with detailed measures of the environment at the time of birth and the socioeconomic position of the household as captured by the census to estimate the effects of early life circumstances on both general and cause-specific mortality both over time and for specific cohorts. We also estimate the relationship between parental cause of death and child cause of death over time. A key strength to our approach to estimating the intergenerational relationship is that we can explicitly control for occupational changes.

Our choice of the Carolinas has distinct advantages. The Carolinas is one of the few locations that records occupation on death certificates over the entire twentieth century. Even for deceased who are retired or not in the labor force at the time of death, these death certificates list the "usual" occupation, and for some years the education level is also listed. Another advantage is that the Carolinas give us a large baseline of black deaths for the entire period, which will be difficult to obtain for many Northern states throughout the twentieth century. Previous studies of racial mortality gaps that have used a linking strategy have generally neglected Southern states, with only one Southern state (Alabama) selected [Preston et al 1996].

There are disadvantages to the Carolina data as well. While we can capture interstate Carolina migration (the dominant migratory pattern early in the century) we cannot capture the effects of selective migration out of the Carolinas or the South more generally. This would be particularly pressing for the study of mortality by race if black (white) migrants were in better (worse) health than stayers. This is exacerbated by the large migratory flows from the South during the Great Migration. This would suggest that if we find significant black/white mortality differences they could be partially ascribed to selective migration, although our general approach of analyzing pro-

portional changes in mortality over time is robust to this criticism. While there certainly was selective migration– migrants have been shown more likely to be urban and more educated in a variety of studies, it does not appear that migration was related to longer life– there is no statistical difference in the mortality of black migrants versus non-migrants during the Great Migration for cohorts born 1905-1925, either overall or for age specific mortality [Sanders and Muszynska 2009]. Recent work has also documented that blacks migrating out of the South during the Great Migration had *worse* socioeconomic outcomes than those that stayed within the South, counter to the conventional wisdom that migration was beneficial to black migrants [Eichenlaub et al 2010]. Overall, while selection is a concern, there is little evidence that it will impact our estimates with regards to mortality or the proportional changes thereof.

In what follows we describe the data and linking strategy that we used to construct our sample. We then describe the general methodology with regards to cause specific mortality and the effects of early life circumstances on mortality. We then motivate and derive our methodology to estimate the degree of intergenerational links between the mortality of parents and children. In the section that follows we present our results. We conclude by discussing the future data collection efforts, limitations to our results, and future analysis that can be carried out with the data.