Black-White Disparities in Life Expectancy: How Much Can Standard SES Variables Explain?

Michael Geruso*

This version: Dec, 2011. Forthcoming in *Demography*

Abstract

This article quantifies the extent to which socioeconomic and demographic characteristics can account for black-white disparities in life expectancy in the United States. Although many studies have investigated the linkages between race, socioeconomic status, and mortality, this article is the first to measure how much of the life expectancy gap remains after differences in mortality are purged of the compositional differences in socioeconomic characteristics between blacks and whites. The decomposition is facilitated by a reweighting technique that creates counterfactual estimation samples in which the distribution of income, education, employment and occupation, marital status, and other theoretically relevant variables among blacks is made to match the distribution of these variables among whites. For males, 80% of the black-white gap in life expectancy at age 1 can be accounted for by differences in socioeconomic and demographic characteristics. For females, 70% percent of the gap is accounted for. Labor force participation, occupation, and (among women only) marital status have almost no additional power to explain the black-white disparity in life expectancy after precise measures for income and education are controlled for.

Keywords: health disparities, racial disparities, decomposition, health gradient, reweight

* Department of Economics, Princeton University, 348 Wallace Hall, Princeton, NJ 08544 e-mail: mgeruso@princeton.edu

Introduction

In the United States, whites live longer than blacks. Since 1980, the difference in life expectancy at birth between whites and blacks has ranged from 5.1 to 7.1 years (Arias et al. 2010). Mortality differences across racial lines have attracted significant public policy interest in the past two decades, which has been met with a surge of scholarly research on racial disparities in health (for an overview, see Harris 2010). Much of this scholarly work has focused on the role of socioeconomic status (SES) in explaining racial differences in health outcomes, and numerous studies have attempted to decompose racial mortality differences, usually between blacks and whites, into a component that is attributable to group differences in SES. One limitation of past studies is that by focusing on mortality hazards estimated over limited age ranges, these studies have left largely unexamined the extent to which SES explains disparities in life expectancy.

This article quantifies the proportion of the racial gap in life expectancy that can be accounted for by differences in characteristics such as income, education, employment and occupation, homeownership, urban residence, and marital status. It employs a reweighting technique that facilitates the decomposition of mortality differences into a component that reflects compositional differences in socioeconomic and demographic characteristics between blacks and whites.

The technique is compatible with a focus on age-specific mortality rates, excess mortality, or any other measure of mortality, but because the data used here cover almost the entire age distribution, it affords a unique opportunity to investigate how SES differences can account for mortality differences over the life course, as summarized by life expectancy. This is the first study able to offer a precise quantitative result regarding the proportion of the black-white difference in life expectancy that can be accounted for by differences in individual-level

SES variables. Using data from the National Longitudinal Mortality Study (NLMS)—a large, nationally-representative data set with rich information on SES—I find that for males, about 80% of the black-white gap in life expectancy at age 1 can be accounted for by differences in socioeconomic and demographic characteristics; 70% of the gap between black and white females is explained. I also find that labor force participation; occupation in a white-collar job; and, among women only, marital status have no power to explain the black-white life expectancy gap after precise measures for income and education are controlled for.

The hypothesis that racial differences in socioeconomic characteristics can account for some or all of the mortality differences between blacks and whites is quite natural given that income (Case et al. 2002; Case and Paxson 2010; Ecob and Smith 1999; Elo and Preston 1996; Rogot et al. 1992; Sorlie et al. 1995; Warner and Hayward 2006; Winkleby and Cubbin 2003); wealth (Warner and Hayward 2006); education (Elo and Preston 1996; Kimbro et al. 2008; Meara et al. 2008; Rogot et al. 1992; Sorlie et al. 1995; Warner and Hayward 2006; Winkleby and Cubbin 2003); employment and occupation (Rogot et al. 1992; Johnson et al. 1999; Sorlie et al. 1995; Warner and Hayward 2006; Winkleby and Cubbin 2003); marital status (Elo and Preston 1996; Hu and Goldman 1990; Johnson et al. 2000; Sorlie et al. 1995; Warner and Hayward 2006); and, to a lesser extent, urban residence (Elo and Preston 1996; Hayward et al. 1997) have each been shown to be predictive of health outcomes, and blacks systematically attain levels of these variables that would imply a health disadvantage relative to whites.

In two widely cited reviews, Williams and Collins (1995) and Hummer (1996) summarized the literature on race, SES, and health as indicating that racial differences in socioeconomic characteristics can explain much of the black-white mortality gap. The individual studies cited in these reviews (e.g., Rogers 1992), as well as the more recent studies that have followed them (e.g., Elo and Preston 1996; Hayward et al. 1997; Howard et al. 2000; Richardus and Kunst 2001; Warner and Hayward 2006; Winkleby and Cubbin 2003), generally estimate hazards of dying that include an indicator for race. Accounting for the role of SES is accomplished by comparing the magnitude and significance of the race coefficient before and after the inclusion of SES variables. These studies generally estimate hazards over a limited age range, and sometimes for only particular causes of death. The focus of these analyses is justified both by the specific aims of the research and by data limitations because many data sets cover only adult deaths in some age range. However, a natural consequence of this narrow focus is that the interested reader would find it difficult or impossible to derive from the existing body of research an answer to the question of how much of the black-white gap in life expectancy could be attributable to group differences in socioeconomic and demographic characteristics. This is in spite of the fact that almost all studies examining racial mortality differences have motivated their analyses by pointing to racial gaps in life expectancies.

The intuition behind the decomposition methodology used here is to reweight individual respondents in the black subsample according to their SES characteristics so that in aggregate, the reweighted black sample matches the original white sample in terms of relevant characteristics. To take a simple example, if the proportion of low-income respondents in the white sample is 33% and the proportion of low-income respondents in the black sample is 50%, then the reweight would scale down the contribution of low-income blacks to black mortality rates by a factor of 2/3 and scale up the contribution of high-income blacks by a factor of 4/3. This would result in a “counterfactual” black sample that resembles the actual white sample in that 33% of its members were low income. Mortality in the counterfactual black sample can then be compared with the actual white mortality.

The idea of decomposing compositional differences between groups by reweighting is not new (see, e.g., DiNardo et al. 1996). This is, however, the first time that this decomposition strategy has been applied to racial differences in life expectancy. The technique is useful because life expectancy is not an individual-level outcome; therefore, it is not possible to treat it as a left hand–side variable in a regression on individual-level characteristics.[1]

Throughout the analysis, I mechanically treat the SES and demographic variables as explanatory factors for life expectancy. This approach follows much of the literature, but I view it as generating results that form an upper bound on the causal importance of these variables in explaining mortality because it is likely that causality works both from SES to health and from health to SES (Deaton 2002). Although investigation into the causal mechanisms working from SES to health are of paramount importance for developing policy responses to racial disparities, the focus here is not on mechanisms but rather on a careful statistical accounting. I return to this issue in the Discussion section.

The remainder of this article is organized as follows. The next section describes in detail the NLMS. The Methods section explains and formalizes the reweighting technique. Finally, the main results, which take the form of counterfactual black life expectancies, are reported and discussed, and several robustness checks are presented.

Data

[1] Potter (1991) regressed racial differences in life expectancy across metropolitan statistical areas (MSAs) on MSA characteristics. Importantly, whereas Potter explored the effect of area-level socioeconomic characteristics, such as the local employment and poverty rates, the data and technique used here permit attribution of life expectancy differentials to *individual*-level characteristics, including own employment status and own income.

This article uses data from Release 3 of the NLMS public-use file. The NLMS is a prospective mortality study of the noninstitutionalized U.S. population. It links demographic and socioeconomic data from Current Population Surveys (CPS) with death records. The CPS, which forms the baseline of the study, is a nationally representative household survey administered monthly by the U.S. Census Bureau and the Bureau of Labor Statistics. The emphasis of the CPS is on labor force participation and employment, but the survey also collects data on income, education, and a variety of household demographics. Mortality follow-up for 11 years following each baseline CPS is accomplished by matching CPS records to death certificates through the National Death Index (NDI), which is an electronic census of all U.S. death certificates since 1979. Eleven CPS cohorts from the early 1980s constitute the NLMS public-use file. These cohorts are aggregated, and balancing weights are applied so that the aggregated sample reflects the noninstitutionalized U.S. population as of April 1, 1983.

In this study, I compare mortality by sex between non-Hispanic whites and non-Hispanic blacks. From the pool of white and black respondents with a valid survey weight, I eliminate records with missing household income or with missing classification as urban/non-urban. I also eliminate records for individuals aged 25 and older with missing marital status, working status, or educational attainment. These restrictions reduce the sample size by less than 0.02%. The resulting sample covers 869,084 individuals and more than 9 million life-years of follow-up. Summary statistics by race and sex at the individual level are provided in Table 1.

[insert Table 1 around here]

The summary statistics show that incomes are lower among blacks than among whites. Black households are larger, which magnifies differences in household income. Black respondents are less likely to own their residence and more likely to live in an urban neighborhood. Homeownership is reported as a potential proxy for wealth because no direct measure of wealth exists in the NLMS.

Among adults aged 25 and older, blacks have lower educational achievement; in particular, white men in the sample are 2.4 times as likely as black men to earn a college degree or higher. Black men (but not women) are less likely to be in the labor force compared with whites; and conditional on labor force participation, black men and women are both somewhat less likely to be employed and much less likely to be working in a white-collar occupation. Marriage rates among black respondents are significantly lower as well, especially for black women.[2]

In the decomposition that follows, the variables used to calculate the reweighting functions include family income, education, current marital status, homeownership, labor force participation, white-collar occupation, and urban residence. Family income is deflated to 1990 dollars and then reexpressed as a percentage of the 1990 poverty line, where the poverty line is specific to a particular household makeup. This allows the income measure to account for differences in household composition, which is important in this context because blacks are more likely to live in larger households. This relative-to-poverty-line income, which is coded categorically as percentiles in the NLMS data set, enters the analysis as nine income levels, roughly corresponding to deciles. Education is categorized in the same six groups as reported in Table 1. Homeownership, urban residence, current marital status, and labor force participation are all binary variables. White-collar status takes three values: white collar, blue collar, and

[2] The sex difference in current marital status is due entirely to older respondents. Because women live longer than men, they are more likely to be (unmarried) widows and skew the sex balance of marriage rates among both whites and blacks.

missing (to mechanically allow for respondents who are out of the labor force). Family income, homeownership, and urban residence are common to all family members in the household survey and are included in the reweight for all age groups. Education, marital status, and labor market activities are individual attributes included in the reweight for adults aged 25 and older.

An important limitation of the NLMS is that neonatal death is largely missed. Because the baseline survey is derived from the household roster of the CPS, children dying within a short window after birth, possibly in the hospital, have a low probability of being enumerated at baseline. This could lead to downward-biased estimates of mortality at age 0. For this reason, in all the analyses that follow, I present results on life expectancy at age 1 rather than age 0. I discuss the likely consequences of ignoring infant mortality in the Discussion section.

Because the object of interest in this article is a comparison of life expectancy across race, it is important to confirm that life expectancy estimates derived from NLMS replicate the racial disparities found in the standard estimates of life expectancy, such as those published by the National Center for Health Statistics (NCHS). There are several reasons that estimates from the NLMS and NCHS should not match exactly. For one, the NLMS covers only the noninstitutionalized U.S. population. More importantly, the NLMS covers the mortality experience of CPS cohorts spanning a decade, each with a mortality follow-up of 11 years. Pooling across CPS cohorts is useful for creating a sample large enough to support the analysis, but the pooled estimates will not be directly comparable with any yearly estimate from the NCHS.

Nonetheless, to investigate whether the NLMS captures the broad patterns of racial disparities evident in other national data, Table 2 lists my estimates of life expectancy at age 1

(e_1) derived from the NLMS sample and compares these with estimates from the 1987 and 1988 NCHS (NCHS 1990, 1991).

[insert Table 2 around here]

Estimates of e_1 in the NLMS are calculated throughout this article according to standard life table procedures but without making functional form assumptions on the survival curve. Age-specific mortality rates are estimated in one-year intervals up to age 85. For the open-ended age interval beginning at 85, calculations follow standard practice (Preston et al. 2001:48). Because the NLMS public-use file is composed of CPS cohorts from the early 1980s with a mortality follow-up period of 11 years, I include NCHS estimates from 1987 and 1988 for comparison. Of course, there is no reason to expect the pooled NLMS estimates to exactly match the single-year NCHS estimates, but patterns of mortality by race and sex found in the NLMS are broadly consistent with the NCHS.[3] The racial mortality gap is slightly smaller in the NLMS estimates, which estimate black life expectancy at age 1 to be higher than in the NCHS estimates. However, the NCHS numbers reflect the uncommonly bad mortality experience among blacks in the late 1980s, which generated period life expectancies among blacks lower than in the years both preceding and following (Arias et al. 2010).

Methods

Decomposition Strategy

---

[3] Because the NLMS excludes institutionalized persons, it is somewhat surprising that the NLMS estimates of e_1 for whites fall slightly below the NCHS figures. For most ages, age-specific mortality is indeed lower in the NLMS. The lower life expectancy among whites in the NLMS is driven entirely by the age ranges 60 to 85 for men and 65 to 85 for women, over which the white NLMS cohorts experience higher mortality compared to the NCHS.

This reweighting technique used in this article is conceptually related to the decompositions introduced by Blinder (1973) and Oaxaca (1973) to evaluate whether differences in group characteristics, such as education and work experience, could account for the observed differences in wages between men and women or whites and blacks. The Blinder-Oaxaca decompositions relied on the assumption of a linear functional relationship between the independent and dependent variables at the individual level. A direct application of the Blinder-Oaxaca decomposition would be impossible in this context because e_1 is not an individual-level outcome, but it is straightforward to construct a generalized decomposition in the same spirit as Blinder-Oaxaca. Denote white and black life expectancy at age 1 by e_1^{white} and e_1^{black}, respectively. By adding and subtracting a term, one can write the difference in life expectancy as

$$e_1^{\text{white}} - e_1^{\text{black}} = \left(e_1^{\text{white}} - e_1^{\text{black,CF}}\right) + \left(e_1^{\text{black,CF}} - e_1^{\text{black}}\right), \tag{1}$$

where $e_1^{\text{black,CF}}$ is a counterfactual life expectancy among a sample of blacks with characteristics that match the white sample. The second difference contained in parentheses in Eq. (1) is often referred to as the part of the difference in the outcome (here $e_1^{\text{white}} - e_1^{\text{black}}$) that is due to differences in characteristics. Analyzing the fraction of the total gap in life expectancy that is due to differences in characteristics is the focus of this article.

Using the demographic and socioeconomic variables from the baseline CPS surveys, I calculate a variety of counterfactuals in which the black subsample is reweighted to match the white subsample along one or several dimensions of observable characteristics. The details of constructing $e_1^{\text{black,CF}}$ are presented in the next subsection. To quantify the proportion of the black-white mortality gap that can be accounted for by differences in characteristics, I report the counterfactual life expectancies as well as the fraction of the gap explained:

$$\% \text{ Explained} = \frac{e_1^{\text{black,CF}} - e_1^{\text{black}}}{e_1^{\text{white}} - e_1^{\text{black}}}, \tag{2}$$

where the numerator measures difference between the counterfactual and true life expectancies for blacks, and the denominator measures the difference in (true) life expectancy by race. True life expectancies in Eq. (2) are calculated as those reported in Table 2. *% Explained* is intended in the accounting sense, rather than as a strong claim about causality.

Generating Reweighting Functions and Constructing Counterfactuals

Constructing counterfactual life expectancies ($e_1^{\text{black,CF}}$) here involves introducing a reweighting term that multiplies each observation's contribution to the calculation of age-specific mortality rates (ASMRs). Denote the probability density of observable characteristics in each of the black and white subsamples as $f(\mathbf{y}|R = b)$ and $f(\mathbf{y}|R = w)$, respectively, where R specifies race; and $\mathbf{y}$ is a vector of observables, such as income, education, and marital status. In this application, f is always a probability mass function, rather than a continuous density, because the NLMS covariates are coded discretely. Nonetheless, the results are general to both discrete and continuous covariates. The reweighting function $\psi(\mathbf{y})$ is defined as a function that multiplies $f(\mathbf{y}|R = b)$ so that the product equals $f(\mathbf{y}|R = w)$ at each value $\mathbf{y}$. In other words, $\psi(\mathbf{y})$ forces the distribution of characteristics to match some reference distribution: in this case, whites. For illustration, I will derive $\psi(\mathbf{y})$ for the case of matching the income distribution of blacks to whites. Additional reweighting functions are derived in the same manner, and the formulas for each are explicitly listed in the tabulations of results.

Let i and a represent income and age, and let $\mathbf{y}'$ represent the elements of $\mathbf{y}$ other than i and a. Consider the counterfactual in which the black income distribution matches the white income distribution at each age, but the joint conditional distribution of other attributes $f(\mathbf{y}'|i, a, R$

= b) remains the same as in the original black sample. The counterfactual density of attributes in this case can be written as the product of several observable conditional density terms: $f(\mathbf{y}'|i, a, R = b)f(i|a, R = w)f(a|R = b)$. It is useful to construct a reweighting function as follows:

$$f\left(\mathbf{y}'\middle|i,a,R=b\right)f\left(i\middle|a,R=w\right)f\left(a\middle|R=b\right)$$

$$= f\left(\mathbf{y}'\middle|i,a,R=b\right)\psi_{i|a}\left(i,a\right)f\left(i\middle|a,R=b\right)f\left(a\middle|R=b\right) \tag{3}$$

$$= f\left(i,a,\mathbf{y}'\middle|R=b\right)\psi_{i|a}\left(i,a\right), \tag{4}$$

where the reweighting function $\psi_{i|a}(i,a)$ is defined as

$$\psi_{i|a}\left(i,a\right) \equiv \frac{f\left(i\middle|a,R=w\right)}{f\left(i\middle|a,R=b\right)}. \tag{5}$$

Equation (4) shows that the desired counterfactual can be achieved by simply multiplying the true distribution of characteristics by an easily calculable reweighting function. The reweighting function, Eq. (5), is just the density of income (conditional on age) among whites divided by the same density among blacks. Given the income distribution shown in Table 1, this implies that for high levels of income, the numerator in Eq. (5) will be larger than the denominator because relatively more whites than blacks are found at higher income levels. This results in a reweighting function that is greater than 1 at high incomes, up-weighting observations of wealthier blacks within each age band.

This reweighting function can be directly applied to the calculation of ASMRs in order to generate counterfactual mortality rates and, in turn, a counterfactual life expectancy. The counterfactual ASMRs, estimated in one-year bins and denoted ${}_1m_x$ are

$${}_1m_x = \frac{\sum_{j\in J} D_j(x)\cdot wt_j \cdot \psi_{i|a}(i,a)}{\sum_{j\in J} L_j(x)\cdot wt_j \cdot \psi_{i|a}(i,a)}, \tag{6}$$

where subscript j indexes individuals; $L_j(x)$ and $D_j(x)$ are defined as in life tables; $D_j(x)$ serves as an indicator for death at age-last-birthday x during the follow-up period; and $L_j(x)$ represents the fraction of a year, if any, lived in the age interval $(x,x + 1)$ during the follow-up period. The weight, wt_j, is the sample weight original to the NLMS data set. Except for the reweighting function $\psi(\cdot)$, Eq. (6) is just a standard formula for an ASMR.

For every counterfactual considered, reweighting functions are specific to a five-year age group. This age grouping has nothing to do with the calculation of age-specific mortality rates, which are done in one-year bins. Instead, this grouping is necessary to ensure, for example, that a wealthy younger black respondent is not up-weighted to balance a wealthy older white respondent.

Calculating the reweighting functions in practice requires that the actual probability mass function for blacks includes full support over the range of characteristics for which the white probability mass function has support. In other words, to reweight observations in the black subsample that possess some specific combination of socioeconomic attributes, it must be the case that at least some black respondents with that particular combination of attributes are observed. For a few (of the thousands) of cells defined by unique combinations of income, education, marital status, and other attributes, no black respondents are observed. Mathematically, such empty cells would correspond to a 0 in the denominator of Eq. (5).

To address this difficulty, in the analysis that follows, I drop white observations for which no black match is observed—about 3% of the white sample.[4] This creates a selected sample of whites and slightly lowers life expectancy in the restricted white sample. Among white

[4] The reverse scenario, of no support in the white density function but positive support in the black density function, poses no technical problem. It results in a reweighting term equal to 0.

men, the estimate of life expectancy at age 1 drops from 71.3 to 71.2 after removing such cases. Among white women, it declines from 78.1 to 77.9.

To keep comparisons consistent, in the analysis that follows, I always compare the original and counterfactual black life expectancies with these revised white estimates. An alternative approach to addressing the lack of joint support in the probability mass functions of characteristics would be to make the grid of characteristics less fine, such as by lumping more education or income groups together. However, to do so would be to impose more parametric assumptions on the functional relationship between socioeconomic variables and mortality, which would affect results in a far less transparent way than dropping some white observations and reporting the effect on white life expectancy. Nonetheless, as a robustness check I also report results of an alternative scheme that reweights whites rather than blacks. That method drops almost no observations and yields results that closely mirror the main results.

Relationship to Other Demographic Decompositions

An important class of demographic techniques, formalized by Kitagawa (1955), decomposes differences in crude (or total) rates between groups into a component that reflects a compositional difference and a component that reflects differences in specific rates. The prototypical example in Kitagawa (1955) is the decomposition of a difference in crude death rates into a part attributable to differences in the age distributions of the populations and a part attributable to differences in age-specific death rates. Variants on the technique differ in technical details but share the aim of purging or measuring the compositional effects.

The decomposition here is similar in that it evaluates a certain type of compositional effect in accounting for differences in group mortality. Here, the compositional effect to be measured and removed is that of SES differences between race groups. Just as with most standardization techniques, the reweight here permits any distribution (of SES variables) to be used as the "standard" or "reference" distribution. The analysis that follows primarily uses the white sample as the reference population, but also presents results in which the black sample is used as the reference.

Compared with Kitagawa (1955) and later work that builds on it, including Das Gupta (1993), there are two important differences in the decomposition technique described here. The first is that this reweight is applied to individual observations rather than to rates or other statistics.

A second difference and a subtle advantage of the reweighting technique presented here is that the counterfactual generated by an individual-level reweight can be used to calculate any group statistic, including the median and variance, and is not limited to the standardization or decomposition of means, as in Kitagawa (1955), Das Gupta (1993), and the Blinder-Oaxaca method. For a more technical exposition of the reweight details, see DiNardo et al. (1996).

A separate class of demographic decompositions operates on life table quantities to attribute differences in life expectancy between two populations to differences in age-specific mortality rates (Arriaga 1984) or to differences in cause-specific mortality rates (Arriaga 1989). This article is related to, but distinct from, the strand of research on the black-white mortality gap that has focused on such mechanical decompositions. Keith and Smith (1988) decomposed racial differences in life expectancy by cause of death to identify the leading cause contributors. More recently, Harper et al. (2007) reported on the age groups and causes of death that contributed most substantially to changes in the black-white life expectancy gap during the 1980s and 1990s.

Reweighting is a natural complement to such decompositions. After counterfactual black or white samples are created through the individual-level reweight and counterfactual life tables are created, any standard life table technique can be employed. For instance, as a supplementary analysis, I follow Arriaga (1984) to evaluate which age ranges are mechanically responsible for the increase in life expectancy after the black sample has been purged of compositional effects of SES via the reweight. Thus, the present reweighting methodology augments long-established life table decomposition techniques.

Results

Main Results

This section quantifies the extent to which controlling for socioeconomic and demographic covariates via a reweighting function can close the racial mortality gap, measured here as life expectancy at age 1. Tables 3 and 4 report the main results of the article for men and women, respectively. Column 1in each table presents the sample estimate of life expectancy at age 1. This column contains the true sample estimate of white and black life expectancy in the first two rows, followed by various counterfactual estimates of black life expectancy. The set of covariates on which the reweight is performed are displayed to the left. Standard errors are in column 2. The third column computes the level difference of life expectancy between the white sample and the counterfactual black sample, and the fourth column displays the percentage of the gap $\left(e_1^{\text{white}} - e_1^{\text{black}}\right)$ explained, defined as in Eq. (2). Here, 0% explained would correspond to a black life expectancy that did not change from its baseline after reweighting the black sample, whereas 100% explained would mean that black life expectancy converged to white life expectancy after matching characteristics. The final column displays the reweighting function numerator for reference.

[insert Tables 3 and 4 around here]

Because the reweighting methodology used here does not allow the predictive power of a set of covariates to be divided into the contribution of each covariate separately, Tables 3 and 4 present a large set of counterfactual black life expectancies in order to illuminate the relative importance of each covariate. The first set of reweights creates counterfactuals that match on only a single characteristic. Then multiple characteristics are included simultaneously so that the black sample is made to match the distribution of characteristics in the white sample along two, three, or four dimensions. By scanning down the table, the reader can judge the relative importance of various characteristics by observing how the inclusion of an additional reweighting variable affects the total predictive power of the set. For instance, among men, income alone accounts for 52% of the gap, and education alone accounts for 26%, but income and education (which are positively correlated) together account for 62% of the racial gap in life expectancy.

For both men and women, income is the most important sole predictor of life expectancy, accounting for 52% and 59% of the gap among males and females, respectively. Education and marital status are also important standalone predictors among both sexes, ranging in predictive power from 20% to 29%. To a lesser extent, homeownership (which may be serving as a proxy for unobserved wealth) is predictive, explaining 16% and 23% among males and females, respectively. Urban residence seems to have almost no predictive power by itself, and men and women differ in the extent to which labor market activities can account for any part of the gap. Recall that unlike black men, black women are *more* likely than their white counterparts to be participating in the labor force, so it is unsurprising that the effect of reweighting labor market activity would not be symmetrical across gender. Moreover, labor market activity and occupation covary strongly with income and education, making it most appropriate to consider the contribution of these characteristics jointly with income, and perhaps with education. Although reweights for single variables are useful for gauging the relative importance of each variable in a set, it is most informative to consider the maximum predictive power when multiple dimensions of characteristics are matched simultaneously, as is done toward the bottom of each table.

One striking difference between the results for males and females is the contribution of marital status, after income and education are controlled for. Among men, additionally controlling for marital status after controlling for income and education yields a 13 percentage point (75% – 62%) increase in the joint predictive power of the variables. Among women, the additional predictive power of marital status conditional on income and education is 0. This does not imply that marriage is an unimportant explanatory factor, given that its individual contribution in a reweight on income, education, and marital status cannot be separately identified.

There are two reasons, in principle, that marital status could have no additional predictive power among women in this case: either there is no mortality gradient in marital status among black women conditional on income and education, or the process of reweighting black women to match white women along education and income creates a reweighted black sample that incidentally also matches the white sample in terms of the proportion who are married so that there is no compositional difference in marriage left to remove. Here, a lack of a mortality gradient in marriage (conditional on income and education) is, in fact, driving the result because even after reweighting for income and education, marriage rates among black women are still 15 percentage points lower than among white women. The difference across gender is consistent with Hu and Goldman (1990), who found a stronger effect of marriage for men than women and with Waldron et al. (1996), who found that the health benefits of marriage among women are stronger for white than black women and operate partly through an income channel.[5]

Jointly controlling for income, education, and marital status together close fully three-quarters of the life expectancy gap for males, and just less than two-thirds of the gap for females. Urban status, which was not predictive when considered alone, becomes important after income and education are controlled for, explaining an additional 14 percentage points (76% – 62%) among men and 6 percentage points (69% – 63%) among women. Reweighting income, education, and urban residence simultaneously has the effect of scaling up the proportion of urban rich blacks and scaling down the proportion of urban poor blacks to better match the geographic distribution of income among whites. Because mortality improves, this suggests that living in an urban area is associated with a health benefit among higher-income blacks. There is no evidence, however, that urban residence confers any benefit for lower-income blacks. The results accord with Hayward et al. (1997), who found an urban mortality advantage accruing only to high-SES types.

After controlling for income and education, additionally controlling for labor force participation or white-collar status appears to have little effect for men, indicating that income and education are sufficient to capture almost all the predictive power of labor participation and white-collar status with respect to mortality.

---

[5] To the extent that marriage provides benefits to women primarily through income, but to men through other channels such as social ties and relations, it is sensible that the residual predictive power of marital status conditional on income (but with no control for social networks) is much smaller for women than for men.

Among women, the percentage of the gap explained decreases from a reweight on income alone to a reweight on the joint distribution of income, education, and labor force participation. This is because when the black sample is reweighted to match only the income distribution of whites, it creates a counterfactual in which black women are significantly more likely to be employed than white women. After matching on income and education, additionally matching the black sample along the dimension of labor participation results in a counterfactual sample in which fewer black women are employed compared with the income-only reweight; and because employment is associated with a mortality advantage in these data (conditional on income and education), black female life expectancy falls when employed black women are down-weighted. Thus, *% explained* decreases.[6] The result illustrates that decomposition procedures of this type generally impose no mechanical requirement that the percentage explained increases as more variables are added.[7]

The last two rows of Tables 3 and 4 summarize the maximum power of the control variables to account for racial differences in life expectancy at age 1. The predictive power tops off at less than 80% for men and 70% for women after jointly reweighting for income, education, marital status, and either homeownership or urban residence. Labor force participation and

---

[6] A similar mechanism drives the results (among women) when white-collar status is included in the reweight after conditioning on income and education.

[7] This feature is shared with other decomposition and standardization techniques. As Kitagawa (1955:1148) explained, "The difference between two crude rates is not the equivalent of a concept like total variance of a dependent variable in regression analysis, for example, which will be increasingly 'explained' as more independent variables are added to the regression equation."

white-collar occupation status are left out because they have little predictive power after including income and education.

These last two rows, which match blacks to whites along four dimensions of characteristics, differ slightly from the other counterfactuals in the table because additional observations from the white sample must be dropped to maintain common support over the finer grid of characteristics. This results in a slightly different estimation sample for calculating white (but not black) life expectancy, which is an input to columns 3 and 4 of the tables. The actual values of e_1^{white} used in these two rows in each table are reported in the table notes and differ trivially from the baseline e_1^{white} in most cases.

It would be desirable to include both urban residence and homeownership simultaneously with income, education, and marriage. However, reweighting jointly along all five dimensions would result in many points at which the probability mass functions for blacks and whites lack common support. For adults alone, calculating the last row in Tables 3 and 4 without dropping observations requires observing black respondents in 2,592 cells (defined by 9 income groups $\times$ 6 education groups $\times$ 2 marital statuses $\times$ 2 urban residence statuses $\times$ 12 five-year age groups). It is possible, though, to trade off the number of variables against how finely each is coded. Therefore, I recode income into five levels (from nine) and jointly reweight on income, education, marital status, homeownership, and urban residence within each five-year age group. The results, not listed in a table, are consistent with Tables 3 and 4: 78% of the racial gap among males and 71% of the racial gap among females is explained.

It is straightforward to evaluate which age-specific mortality rates contribute most to the life expectancy gains resulting when the black sample is reweighted to match the white sample in socioeconomic and demographic characteristics. Applying Arriaga's (1984) technique to compare the life tables of the actual and counterfactual black samples, I find that the age-specific mortality rates mechanically responsible for the life expectancy gains under the counterfactual are concentrated primarily in the age range 40 to 69. For both men and women, more than 70% of the gains originate in this range (for a reweight on income, education, and marriage).

In Tables 6 and 7 in the appendix, I present the results of an exercise similar to the main analysis of Tables 3 and 4, but for which the characteristics in the white sample are made to match the characteristics in the black sample. Reweighting the white sample, in contrast with reweighting the black sample, has the feature that almost no observations need be dropped (less than 0.005%). Because white respondents outnumber black respondents roughly eight to one in the baseline survey, it is rare that a cell defined by a combination of characteristics contains a black, but not a white, respondent. Results for the white reweight are similar, but the interpretation is slightly different in that it is the relationship between SES and mortality among whites, rather than blacks, that generates the life expectancy counterfactuals.

Finally, because it may be of special interest to focus separately on adult mortality, I have included additional tables in Online Resource 1 that repeat the accounting in Tables 3 and 4, but for life expectancy at age 30 rather than age 1. These results, as well, are broadly similar to the main analysis.

Robustness

A potential measurement concern involves the sensitivity of the results to the way in which individual variables are coded. To illustrate how the predictive power of SES can depend on the degree of flexibility in the empirical specification, I take the simple case in which income is recoded into a smaller number of wider bins. Table 5 presents a reweight on income only, which is coded progressively coarser moving vertically down each panel.

[insert Table 5 around here]

Each panel in Table 5 begins by restating the result in which income is categorized into nine groups that align roughly with deciles. The rows below it collapse income into wider bins: first quintiles, then terciles, and finally a single division at the median.[8] For both sexes, the explanatory power of income diminishes somewhat in moving from nine groups to five groups, but moving to three groups and especially to just two groups significantly reduces the power of income to account for mortality differences. Jointly reweighting (or analogously in a regression, interacting) with a flexibly parameterized education variable could ameliorate some of the loss of predictive power because education covaries closely with income and can capture some of the income variation obscured by mismeasurement.

This characterization is relevant to the many studies that are forced by data limitations to use very coarse categorizations of SES. Table 5 suggests that unless the theoretically relevant variables are well measured and enter the empirical specification flexibly (regardless of whether that specification is a hazard regression or a reweight), one can easily mistake a lack of fidelity in the data for a lack of explanatory power of SES. This point about measurement error in explanatory variables is, of course, generally true, but it is particularly relevant in this context where authors are frequently constrained to enter SES as high/medium/low or high/low.

In this article, a flexible specification allows for hundreds and sometimes thousands of SES groups, each defined by a unique combination of covariates. Nonetheless, two important variables—marital status and occupation—are coded dichotomously in the main analysis. This choice of coding was based on preliminary sensitivity analysis. The remainder of this section explores the consequences of allowing both variables to enter the reweight more flexibly.

[8] Percentiles are approximate.

To explore the robustness of the results to an alternative coding of marriage, I rerun the main analysis with marriage entered as five categories: married, divorced, widowed, separated, and never married. Given the mechanics of the decomposition technique, this requires reducing the number of income categories from nine to five to avoid dropping more than a small fraction of the white sample. For both men and women, the decompositions that included the more finely-coded marriage variable were statistically and practically indistinct from those of the main analysis. For men, marriage alone accounted for 25% of the gap (compared with 23% in Table 3 in the main analysis). When entered jointly with income and education, it accounted for 71% (versus 75%). For women, marriage alone accounted for 21% (versus 20% in Table 4) and 62% (versus 63%) when entered jointly with income and education.

A similar robustness check was performed with respect to white-collar status/occupation. Occupation was recoded into eight categories: managerial; professional and technical; sales; support and clerical; service; agriculture; crafts, laborers, and transportation; and missing (for those not in the labor force). Like with marriage, these results were remarkably similar to those in Tables 3 and 4. In no case for either sex was the difference between the robustness check value and the corresponding value from the main tables even marginally statistically significant.

Discussion

These results are not intended to imply that an intervention that managed to close racial gaps in education, income, and marriage rates between blacks and whites should be expected to close mortality gaps by the proportions displayed in the tables. It is doubtful that all the observed association between SES and mortality reflects a causal effect running from SES to mortality. It is widely acknowledged, for instance, that selection into marriage is at least part of the reason for the observed association between marital status and health (see, e.g., Goldman 1993; Lillard and

Panis 1996; Murray 2000). With regard to income, Case and Paxson (2011) demonstrated a channel running from health to workplace promotion and income. And since at least the Farrell and Fuchs (1982) study of education and smoking, there has been strong empirical grounding for doubts as to whether the poor health behaviors that are associated with lower education necessarily reveal anything about the causal impact of schooling on health.

To the extent that part of the correlation between SES and mortality is due to third factors and reverse causality, the decomposition results presented in this article will tend to overestimate causal effects. Thus my preferred interpretation of the results presented in the last section is as an upper bound on the causal impact of the variables considered.

Even under the most generous assumption that all the observed correlation between SES and mortality represents a causal impact of SES on health, around 20% and 30% of the racial mortality gaps among men and women, respectively, remain. The existence of a residual in related studies has been hypothesized to be attributable to causes such as differential access to quality medical care; psychosocial stress due to racial discrimination; and, most commonly, failure to capture the "right" SES variables, such as peer networks or detailed neighborhood characteristics (Williams and Collins 2001; Winkleby and Cubbin 2003). And of course, many important non-SES factors are not captured in this study (except to the extent that they covary with the variables examined). Among these are health behaviors, the most significant of which may be smoking.

This study benefits from a large data set that includes detailed measures of SES and covers mortality over a wide range of ages. Nonetheless, important limitations of the data are apparent. The analysis here has been limited to a comparison of non-Hispanic blacks and whites, both in keeping with previous literature, and due to the practical limitations imposed by small sample sizes for other race groups.

The focus on e_1 necessarily ignores the important contribution of infant mortality to life expectancy at age 0. To put this in perspective, calculations based on NCHS Vital Statistics data from the mid-1980s reveal that mortality between ages 0 and 1 accounted for around 10% of the e_0 gap between white and black males and for around 13% of the e_0 gap between white and black females in this time period. There is some external evidence that the role of SES in accounting for mortality differences between ages 0 and 1 is likely similar to my findings above age 1. For example, Finch (2003), focusing on infant mortality, found that income is the single most important predictor of infant mortality, outstripping parents' education and occupational grade in explanatory power. This suggests that with more complete data, results for a decomposition of e_0 disparities would be similar to results for e_1.

Finally, it is important to delineate the circumstances under which a decomposition of the type found here is preferable to a hazard regression analysis. Hazards are usually the most appropriate choice when the research aim is to quantify the association between a socioeconomic variable and mortality in some age range. However, the sign, size, and significance of the coefficient on an SES variable in a regression is uninformative, in itself, as to that variable's importance in explaining group differences in mortality, which are a function both of the coefficient and of a group difference in the variable that the coefficient multiplies. Thus, the reweight technique is best viewed as an alternative to hazard regressions when the goal is decomposing group differences and individual-level data are available. It can also bypass model estimation, and it allows the decomposition of statistics other than rates and means, such as variance.

Conclusion

This article is intended to offer a straightforward answer to the question of how much of the black-white disparity in life expectancy can be accounted for by group differences in socioeconomic and demographic characteristics. The findings here that differences in such characteristics can account for up to 80% of the gap between white and black males and up to 70% of the gap between white and black females are consistent with an extensive literature in sociology, demography, and epidemiology that finds that SES, most often operationalized as income and education, can account for a large part of the racial disparities in U.S. mortality.

Reweighting and the associated decomposition presented in this article purge mortality differences of the compositional effects of SES. Although the analysis here focuses on life expectancy (with the intention of parsimoniously summarizing the total effects of mortality differentials), the decomposition is compatible with the examination of any mortality measure, including age- and cause-specific rates. Indeed, further decomposing these results into contributions of specific causes of death is an interesting avenue of future research.

Acknowledgments I would like to thank Anne Case, Angus Deaton, Thomas Espenshade, Scott Lynch, John Papp, and seminar participants at Princeton's Center for Health and Wellbeing for helpful comments. I also thank the referees for many useful suggestions. This article uses data supplied by the National Heart, Lung, and Blood Institute; NIH; and DHHS from the National Longitudinal Mortality Study. The views expressed in this article are those of the author and do not necessarily reflect the views of the National Heart, Lung, and Blood Institute; the Census Bureau; or the National Center for Health Statistics.

References

Arias, E., Rostron, B. L., & Tejada-Vera, B. (2010). United States life tables, 2005. *National Vital Statistics Reports*, *58*(10). Hyattsville, MD: National Center for Health Statistics.

Arriaga, E. E. (1984). Measuring and explaining the change in life expectancies. *Demography, 21,* 83–96.

Arriaga, E. E. (1989). Changing trends in mortality decline during the last decades. In L. Ruzicka, G. Wunsch, & P. Kane (Eds.), *Differential mortality: Methodological issues and biosocial factors* (pp. 105–129). Oxford, UK: Clarendon Press.

Blinder, A. S. (1973). Wage discrimination: Reduced form and structural estimates. *Journal of Human Resources*, *8,* 436–455.

Case, A., Lubotsky, D., & Paxson, C. (2002). Economic status and health in childhood: The origins of the gradient. *American Economic Review*, *92,* 1308–1334.

Case, A., & Paxson, C. (2011). “The long reach of childhood health and circumstance: Evidence from the Whitehall II Study.” *The Economic Journal*, *121,* F183–F204.

Chiang, C. L. (1968). *Introduction to stochastic processes in biostatistics*. New York: John Wiley & Sons.

Das Gupta, P. (1993). *Standardization and decomposition of rates: A user's manual*. Washington, DC: U.S. Government Printing Office.

Deaton, A. (2002). Policy implications of the gradient of health and wealth. *Health Affairs, 21*(2), 13–30.

DiNardo, J., Fortin, N. M., & Lemieux, T. (1996). Labor market institutions and the distribution of wages, 1973–1992: A semiparametric approach. *Econometrica, 64,* 1001–1044.

Ecob, R., & Smith, G. D. (1999). Income and health: What is the nature of the relationship? *Social Science & Medicine*, *48,* 693–705.

Elo, I., & Preston, S. (1996). Educational differentials in mortality: United States, 1979–1985. *Social Science & Medicine*, *42*(1):47–57.

Farrell, P., & Fuchs, V. (1982). Schooling and health: The cigarette connection. *Journal of Health Economics, 1*, 217-230.

Finch, B. (2003). Early origins of the gradient: The relationship between socioeconomic status and infant mortality in the United States. *Demography, 40,* 675–699.

Goldman, N. (1993). Marriage selection and mortality patterns: Inferences and fallacies. *Demography, 30,* 189–208.

Harper, S., Lynch, J., Burris, S., & Davey Smith, G. (2007). Trends in the black-white life expectancy gap in the United States, 1983–2003. *Journal of the American Medical Association*, *297*, 1224–1232.

Harris, K. M. (2010). An integrative approach to health. *Demography, 47,* 1–22.

Hayward, M., Pienta, A., & McLaughlin, D. (1997). Inequality in men's mortality: The socioeconomic status gradient and geographic context. *Journal of Health and Social Behavior, 38,* 313–330.

Howard, G., Anderson, R. T., Russell, G., Howard, V. J., & Burke, G. L. (2000). Race, socioeconomic status, and cause-specific mortality. *Annals of Epidemiology, 10,* 214–223.

Hu, Y., & Goldman, N. (1990). Mortality differentials by marital status: An international comparison. *Demography*, *27*, 233–250.

Hummer, R. A. (1996). Black-white differences in health and mortality: A review and conceptual model. *Sociological Quarterly, 37,* 105–125.

Johnson, N. J., Backlund, E., Sorlie, P. D., & Loveless, C. A. (2000). Marital status and mortality: The National Longitudinal Mortality Study. *Annals of Epidemiology, 10,* 224–238.

Johnson, N. J., Sorlie, P. D., & Backlund, E. (1999). The impact of specific occupation on mortality in the U.S. National Longitudinal Mortality Study. *Demography, 36,* 355–367.

Keith, V. M., & Smith, D. P. (1988). The current differential in black and white life expectancy. *Demography, 25,* 625–632.

Kimbro, R. T., Bzostek, S., Goldman, N., & Rodríguez, G. (2008). Race, ethnicity, and the education gradient in health. *Health Affairs, 27*, 361–372.

Kitagawa, E. M. (1955). Components of a difference between two rates. *Journal of the American Statistical Association, 50,* 1168–1194.

Lillard, L. A., & Panis, C. W. A. (1996). Marital status and mortality: The role of health. *Demography, 33,* 313-327.

Meara, E. R., Richards, S., & Cutler, D. M. (2008). The gap gets bigger: Changes in mortality and life expectancy, by education, 1981–2000. *Health Affairs, 27,* 350–360.

Murray, J. E. (2000). Marital protection and marital selection: Evidence from a historical-prospective sample of American men. *Demography, 37,* 511–521.

National Center for Health Statistics. (1990). *Vital statistics of the United States, 1987: Life tables* (Vol. II, Section 6). Washington, DC: Public Health Service.

National Center for Health Statistics. (1991). *Vital statistics of the United States, 1988: Life tables* (Vol. II, Section 6). Washington, DC: Public Health Service.

Oaxaca, R. (1973). Male-female wage differentials in urban labor markets. *International Economic Review, 14*, 693–709.

Potter, L. B. (1991). Socioeconomic determinants of white and black males' life expectancy differentials, 1980. *Demography, 28,* 303–321.

Preston, S., Heuveline, P., & Guillot, M. (2001). *Demography: Measuring and modeling population processes*. Oxford, UK: Wiley-Blackwell.

Richardus, J. H., & Kunst, A. E. (2001). Black-white differences in infectious disease mortality in the United States. *American Journal of Public Health, 91,* 1251–1253.

Rogers, R. G. (1992). Living and dying in the U.S.A.: Sociodemographic determinants of death among blacks and whites. *Demography, 29,* 287–303.

Rogot, E., Sorlie, P. D., & Johnson, N. J. (1992). Life expectancy by employment status, income, and education in the National Longitudinal Mortality Study. *Public Health Reports, 107,* 457–461.

Sorlie, P. D., Backlund, E., & Keller, J. B. (1995). US mortality by economic, demographic, and social characteristics: The National Longitudinal Mortality Study. *American Journal of Public Health, 85,* 949–956.

Waldron, I., Hughes, M. E., & Brooks, T. L. (1996). Marriage protection and marriage selection-prospective evidence for reciprocal effects of marital status and health. *Social Science & Medicine, 43,* 113–123.

Warner, D., & Hayward, M. (2006.) Early-life origins of the race gap in men's mortality. *Journal of Health and Social Behavior, 47,* 209–226.

Williams, D. R., & Collins, C. (1995). US socioeconomic and racial differences in health: Patterns and explanations. *Annual Review of Sociology, 21,* 349–386.

Williams, D. R., & Collins, C. (2001). Racial residential segregation: A fundamental cause of racial disparities in health. *Public Health Reports, 116*, 404–416.

Winkleby, M., & Cubbin, C. (2003). Influence of individual and neighbourhood socioeconomic status on mortality among black, Mexican-American, and white women and men in the United States. *British Medical Journal*, *57*, 444–452.

Table 1 Summary statistics at baseline

	White		Black	
	Male	Female	Male	Female
Panel A: All Respondents				
Persons	372,639	400,209	42,880	53,356
Deaths	40,731	35,470	4,610	4,307
Family income (1990 $)				
<10,000	0.08	0.11	0.26	0.32
10,000–19,999	0.15	0.17	0.25	0.26
20,000–29,999	0.21	0.21	0.21	0.19
30,000–39,999	0.17	0.16	0.12	0.10
40,000–65,999	0.30	0.27	0.14	0.12
≥75,000	0.08	0.07	0.02	0.02
Household size	3.53	3.39	4.16	4.14
Own home	0.75	0.73	0.50	0.47
Urban	0.64	0.66	0.83	0.84
Age	32.58	34.89	27.40	29.70
Panel B: Respondents Aged 25 and Older at Baseline				
Persons	222,590	252,307	20,255	28,607
Education (years)				
Elementary (0–4)	0.02	0.01	0.09	0.05
Middle school (5–8)	0.11	0.11	0.16	0.16
Some high school (9–11)	0.11	0.13	0.19	0.21
High school (12)	0.35	0.44	0.32	0.34
Some college (13–15)	0.17	0.16	0.14	0.14
College and graduate school (16+)	0.24	0.16	0.10	0.09
In labor force	0.80	0.52	0.75	0.58
Employed	0.90	0.89	0.83	0.82
White-collar job	0.49	0.72	0.26	0.49
Currently married	0.78	0.69	0.59	0.43
Age	46.25	48.02	44.66	45.26

Notes: Income deflated to 1990 dollars. Except for the variables *persons*, *deaths*, *household size*, and *age*, the numbers reported are fractions. Employment and white-collar status are calculated conditional on labor force participation.

Table 2 Life expectancy comparability between the NLMS and NCHS

	White		Black	
Life Expectancy at Age 1 (e_1)	Male	Female	Male	Female
NLMS	71.3	78.1	66.6	74.3
NCHS, 1987	71.9	78.5	65.5	73.8
NCHS, 1988	72.0	78.5	65.2	73.6

Sources: NLMS estimates from author's calculations. NCHS estimates are from vital statistics (NCHS 1990, 1991).

Table 3 Decomposition of e_1 differences by SES and demographics: Results for males, reweighting the black sample

	(1) e_1	(2) SE(e_1)	(3) $e_1^{\text{white}} - e_1$	(4) % Explained	(5) Reweight Numerator
White	71.20	0.07	0.00		
Black	66.64	0.28	4.56		
Black Reweight on					
Income (inc)	69.02	0.38	2.17	52	$f(inc\|a, R = w)$
Education (ed)	67.83	0.29	3.37	26	$f(ed\|a, R = w)$
Marital status (ms)	67.71	0.29	3.49	23	$f(ms\|a, R = w)$
Own home (own)	67.37	0.35	3.83	16	$f(own\|a, R = w)$
In labor force (lf)	67.59	0.29	3.61	21	$f(lf\|a, R = w)$
White-collar job (wc)	67.89	0.29	3.31	27	$f(wc\|a, R = w)$
Urban residence (urb)	66.76	0.28	4.44	3	$f(urb\|a, R = w)$
Inc, ed	69.45	0.39	1.75	62	$f(inc, ed\|a, R = w)$
Ed, ms	68.79	0.29	2.41	47	$f(ed, ms\|a, R = w)$
Inc, ms	69.67	0.39	1.53	66	$f(inc, ms\|a, R = w)$
Inc, ed, ms	70.06	0.39	1.14	75	$f(inc, ed, ms\|a, R = w)$
Inc, ed, own	69.82	0.44	1.38	70	$f(inc, ed, own\|a, R = w)$
Inc, ed, lf	69.54	0.39	1.65	64	$f(inc, ed, lf\|a, R = w)$
Inc, ed, wc	69.49	0.39	1.71	62	$f(inc, ed, wc\|a, R = w)$
Inc, ed, urb	70.09	0.40	1.11	76	$f(inc, ed, urb\|a, R = w)$
Inc, ed, ms, own	70.18	0.44	1.10	76	$f(inc, ed, ms, own\|a, R = w)$
Inc, ed, ms, urb	70.19	0.40	1.04	77	$f(inc, ed, ms, urb\|a, R = w)$

Notes: Column 1 gives actual and counterfactual values for life expectancy at age 1. Standard errors in column 2 are calculated following Chiang (1968). Column 3 displays the difference between column 1 and the value of e_1 in the white subsample (e_1^{white}). In all but the last two rows, the value of e_1^{white} used throughout the table to calculate columns 3 and 4 is the value reported in the first row of column 1. The values of e_1^{white} used for calculations in the last two rows are 71.28 and 71.23, respectively. See the text for further details.

Table 4 Decomposition of e_1 differences by SES and demographics: Results for females, reweighting the black sample

	(1) e_1	(2) SE(e_1)	(3) $e_1^{\text{white}} - e_1$	(4) % Explained	(5) Reweight Numerator
White	77.91	0.07	0.00		
Black	74.34	0.23	3.57		
Black Reweight on					
Income (inc)	76.46	0.27	1.44	59	$f(inc\|a, R = w)$
Education (ed)	75.37	0.23	2.54	29	$f(ed\|a, R = w)$
Marital status (ms)	75.07	0.23	2.84	20	$f(ms\|a, R = w)$
Own home (own)	75.15	0.23	2.76	23	$f(own\|a, R = w)$
In labor force (lf)	74.09	0.23	3.82	7	$f(lf\|a, R = w)$
White-collar job (wc)	74.44	0.23	3.47	3	$f(wc\|a, R = w)$
Urban residence (urb)	74.42	0.23	3.49	2	$f(urb\|a, R = w)$
Inc, ed	76.59	0.27	1.32	63	$f(inc, ed\|a, R = w)$
Ed, ms	75.83	0.23	2.08	42	$f(ed, ms\|a, R = w)$
Inc, ms	76.66	0.27	1.25	65	$f(inc, ms\|a, R = w)$
Inc, ed, ms	76.58	0.26	1.33	63	$f(inc, ed, ms\|a, R = w)$
Inc, ed, own	76.73	0.25	1.18	67	$f(inc, ed, own\|a, R = w)$
Inc, ed, lf	76.11	0.27	1.80	49	$f(inc, ed, lf\|a, R = w)$
Inc, ed, wc	76.29	0.27	1.61	55	$f(inc, ed, wc\|a, R = w)$
Inc, ed, urb	76.79	0.28	1.12	69	$f(inc, ed, urb\|a, R = w)$
Inc, ed, ms, own	76.77	0.25	1.25	66	$f(inc, ed, ms, own\|a, R = w)$
Inc, ed, ms, urb	76.69	0.27	1.24	65	$f(inc, ed, ms, urb\|a, R = w)$

Notes: Column 1 gives actual and counterfactual values for life expectancy at age 1.Standard errors in column 2 are calculated following Chiang (1968). Column 3 displays the difference between column 1 and the value of e_1 in the white subsample (e_1^{white}). In all but the last two rows, the value of e_1^{white} used throughout the table to calculate columns 3 and 4 is the value reported in the first row of column 1. The values of e_1^{white} used for calculations in the last two rows are 78.02 and 77.93, respectively. See the text for further details.

Table 5 Sensitivity analysis: Alternative categorizations of income

	e_1	SE(e_1)	$e_1^{\text{white}} - e_1$	% Explained
Panel A: Males				
White	71.20	0.07	0.00	
Black	66.64	0.28	4.56	
Black reweight on				
Income (9 groups)	69.02	0.38	2.17	52
Income (5 groups)	68.98	0.34	2.22	51
Income (3 groups)	68.76	0.32	2.44	46
Income (2 groups)	68.24	0.31	2.96	35
Panel B: Females				
White	77.91	0.07	0.00	
Black	74.34	0.23	3.57	
Black reweight on				
Income (9 groups)	76.46	0.27	1.44	59
Income (5 groups)	76.21	0.24	1.70	52
Income (3 groups)	76.07	0.24	1.83	49
Income (2 groups)	75.62	0.24	2.28	36

Notes: Standard errors are calculated following Chiang (1968).

Appendix

Table 6 Decomposition of e_1 differences by SES and demographics: Results for males, reweighting white sample

	(1) e_1	(2) SE(e_1)	(3) $e_1 - e_1^{\text{black}}$	(4) % Explained	(5) Reweight Numerator
Black	66.64	0.28	0.00		
White	71.31	0.07	4.67		
White Reweight on					
Income (inc)	69.17	0.09	2.53	46	$f(inc \mid a, R = b)$
Education (ed)	69.86	0.07	3.22	31	$f(ed \mid a, R = b)$
Marital status (ms)	70.35	0.07	3.71	21	$f(ms \mid a, R = b)$
Own home (own)	70.49	0.08	3.85	18	$f(own \mid a, R = b)$
In labor force (lf)	70.36	0.07	3.72	20	$f(lf \mid a, R = b)$
White-collar job (wc)	70.02	0.07	3.38	28	$f(wc \mid a, R = b)$
Urban residence (urb)	71.30	0.07	4.66	0	$f(urb \mid a, R = b)$
Inc, ed	68.69	0.09	2.05	56	$f(inc, ed \mid a, R = b)$
Ed, ms	69.01	0.07	2.37	49	$f(ed, ms \mid a, R = b)$
Inc, ms	68.53	0.09	1.89	60	$f(inc, ms \mid a, R = b)$
Inc, ed, ms	68.07	0.09	1.43	69	$f(inc, ed, ms \mid a, R = b)$
Inc, ed, own	68.29	0.08	1.65	65	$f(inc, ed, own \mid a, R = b)$
Inc, ed, lf	68.60	0.09	1.96	58	$f(inc, ed, lf \mid a, R = b)$
Inc, ed, wc	68.60	0.09	1.96	58	$f(inc, ed, wc \mid a, R = b)$
Inc, ed, urb	68.39	0.08	1.75	63	$f(inc, ed, urb \mid a, R = b)$
Inc, ed, ms, own	67.89	0.08	1.25	73	$f(inc, ed, ms, own \mid a, R = b)$
Inc, ed, ms, urb	67.91	0.08	1.27	73	$f(inc, ed, ms, urb \mid a, R = b)$

Notes: Standard errors in column 2 are calculated following Chiang (1968). Throughout the table, the value of e_1^{black} used to calculate columns 3 and 4 is the value reported in the first row of column 1.

Table 7 Decomposition of e_1 differences by SES and demographics: Results for females, reweighting white sample

	(1) e_1	(2) SE(e_1)	(3) $e_1 - e_1^{\text{black}}$	(4) % Explained	(5) Reweight Numerator
Black	74.34	0.23	0.00		
White	78.06	0.07	3.72		
White Reweight on					
Income (inc)	76.39	0.07	2.05	45	$f(inc \mid a, R = b)$
Education (ed)	77.11	0.07	2.77	26	$f(ed \mid a, R = b)$
Marital status (ms)	77.35	0.07	3.01	19	$f(ms \mid a, R = b)$
Own home (own)	77.36	0.07	3.02	19	$f(own \mid a, R = b)$
In labor force (lf)	78.13	0.07	3.79	–2	$f(lf \mid a, R = b)$
White-collar job (wc)	77.99	0.07	3.64	2	$f(wc \mid a, R = b)$
Urban residence (urb)	77.99	0.07	3.65	2	$f(urb \mid a, R = b)$
Inc, ed	76.01	0.07	1.67	55	$f(inc, ed \mid a, R = b)$
Ed, ms	76.45	0.07	2.11	43	$f(ed, ms \mid a, R = b)$
Inc, ms	76.08	0.07	1.74	53	$f(inc, ms \mid a, R = b)$
Inc, ed, ms	75.71	0.07	1.37	63	$f(inc, ed, ms \mid a, R = b)$
Inc, ed, own	75.67	0.07	1.33	64	$f(inc, ed, own \mid a, R = b)$
Inc, ed, lf	76.33	0.07	1.99	47	$f(inc, ed, lf \mid a, R = b)$
Inc, ed, wc	76.27	0.07	1.93	48	$f(inc, ed, wc \mid a, R = b)$
Inc, ed, urb	75.69	0.07	1.34	64	$f(inc, ed, urb \mid a, R = b)$
Inc, ed, ms, own	75.50	0.07	1.16	69	$f(inc, ed, ms, own \mid a, R = b)$
Inc, ed, ms, urb	75.47	0.07	1.13	70	$f(inc, ed, ms, urb \mid a, R = b)$

Notes: Standard errors in column 2 are calculated following Chiang (1968). Throughout the table, the value of e_1^{black} used to calculate columns 3 and 4 is the value reported in the first row of column 1.