

Heterogeneity in Expected Longevities

Josep Pijoan-Mas · José-Víctor Ríos-Rull

Published online: 13 November 2014
© Population Association of America 2014

Abstract We develop a new methodology to compute differences in the expected longevity of individuals of a given cohort who are in different socioeconomic groups at a certain age. We address the two main problems associated with the standard use of life expectancy: (1) that people's socioeconomic characteristics change, and (2) that mortality has decreased over time. Our methodology uncovers substantial heterogeneity in expected longevities, yet much less heterogeneity than what arises from the naive application of life expectancy formulae. We decompose the longevity differences into differences in health at age 50, differences in the evolution of health with age, and differences in mortality conditional on health. Remarkably, education, wealth, and income are health-protecting but have very little impact on two-year mortality rates conditional on health. Married people and nonsmokers, however, benefit directly in their immediate mortality. Finally, we document an increasing time trend of the socioeconomic gradient of longevity in the period 1992–2008, and we predict an increase in the socioeconomic gradient of mortality rates for the coming years.

Keywords Cohort life expectancy · Expected longevity · Health · Socioeconomic gradient

Introduction

The negative relationship between mortality rates and socioeconomic status is well documented. In a seminal work, Kitagawa and Hauser (1973) showed that U.S. mortality rates in 1960 were inversely related to education and income. Since then, a large body of literature has emerged confirming the socioeconomic gradient of

J. Pijoan-Mas (✉)
CEMFI, Casado del Alisal 5, 28014 Madrid, Spain
e-mail: pijoan@cemfi.es

J. Pijoan-Mas
Center for Economic Policy Research, 77 Bastwick Street, London EC1 3PZ, UK

J.-V. Ríos-Rull
University of Minnesota, Federal Reserve Bank of Minneapolis, CAERP, CEPR, and NBER,
Minneapolis, MN, USA

mortality rates, which is found in education and income but also in wealth, labor market occupation, or marital status.¹

Much less is known, however, about how socioeconomic differences in mortality rates aggregate over the life cycle and create differences in the life duration of people of a given cohort. One approach to addressing this issue is to use differences in period life expectancies, which mechanically aggregate socioeconomic differences in the mortality rates of a given calendar year. For instance, Brown (2002) and Meara et al. (2008) computed differences in period life expectancies according to education level, and Lin et al. (2003) and Singh and Siahpush (2006) did so for other measures of socioeconomic status. However, the use of period life expectancies has two significant drawbacks. First, period life expectancy measures do not account for the possibility that socioeconomic characteristics of individuals evolve over time; that is, because membership in a given population subgroup may change over the life cycle, so do the relevant mortality rates. This is typically the case for any measure of socioeconomic status, except for education. Second, period life expectancies typically do not account for the fact that mortality rates tend to decline over time and may decline at different rates for people in different socioeconomic groups. Lee and Carter (1992), as well as others using related methods, have addressed the time-changing problem of mortality rates by estimating an age-specific time component of mortality rates that can be used for extrapolation. However, the time effects on mortality rates may not be independent from the socioeconomic status of individuals. More importantly, exploiting time series variation ignores important current observable information that may have significant predictive power. To sum up, the static picture that emerges from differences in period life expectancy by different socioeconomic groups may not be a good proxy of actual differences in expected longevity within a cohort of individuals.

The first contribution of this article is to develop and implement a new measure of the expected duration of life of a given cohort—or *cohort life expectancy*—that addresses these two problems. We call this measure *expected longevity*, and we document its socioeconomic gradient. In a first step, we exploit the panel structure of the Health and Retirement Study (HRS) to estimate age-specific survival rates conditional on a socioeconomic characteristic of interest z , as well as its age-specific transition probabilities. The transitions for the socioeconomic characteristic z allow us to address the changes in socioeconomic status over the life cycle. The socioeconomic characteristics studied here are education, wealth, nonfinancial income, labor market status, marital status, and smoking. We then link the estimates for all different cohorts in the HRS in order to build expected longevity conditional on the characteristic z at age 50 for individuals born between 1941 and 1954 (who were 50 years of age between 1992 and 2004). Our results uncover substantial heterogeneity in expected longevity across different measures of socioeconomic status, with the socioeconomic gradient being steeper for men than for women. That said, these socioeconomic gradients are 2 to 3 times smaller than those that result from using period life

¹ See, for instance, Montez et al. (2011) and references therein for recent findings of mortality differences by education level. Deaton and Paxson (1994) documented the negative relationship between mortality and family income, after controlling for education. Attanasio and Hoynes (2000) found a negative relationship between mortality and wealth. The Whitehall studies have uncovered important mortality differences according to the employment grade among British civil servants; see, for instance, Marmot et al. (1984, 1991). For mortality rates and marital status, see Hu and Goldman (1990) and the references therein.

expectancies. This confirms the importance of keeping track of the life-cycle evolution of individual socioeconomic characteristics in order to predict the life duration of a given population subgroup.

In a second step, we estimate age-specific survival rates conditional on a socioeconomic characteristic z and on individual (self-assessed) health h and age-specific joint transition probabilities for z and h . The use of information on health h allows us to partly address the changes over time in survival rates and in the transitions for z and to take into account that the transitions on z may have a direct dependence on health. This is quantitatively important: the socioeconomic gradients in expected longevities that we find are larger (between 20 % and almost 100 %) than when we ignore information on health and hence assume that mortality rates and transition functions are constant over time. The reason for this is that the gap in health conditions at age 50 between the most and the least advantaged types has grown over time. Hence, our results point to an increase in the socioeconomic gradient of mortality rates at old ages for the coming years.

The second contribution of this article, closely linked to the first one, is to decompose the socioeconomic gradient of expected longevity into differences in health already present at age 50, changes in health that developed after age 50, and differences in two-year mortality rates that are unrelated to measured health. We find that approximately one-third of the expected longevity differences for education and wealth categories is attributable to health differences at age 50; the remainder is due to the health protection effect of education and wealth in the years after age 50. Interestingly, the effects of education, wealth, and income on two-year mortality rates are very small or null after we control for self-assessed health. This finding is surprising because higher wealth, income, and education suggest a greater ability to pay for medical treatment. Although financial resources may be health-protecting, they do not appear to lower mortality at the onset of terminal diseases, nor do they appear to reduce death-inducing accidents. In contrast, being married and a nonsmoker significantly reduces mortality rates even after we control for differences in measured health. This raises the question of what exactly is behind the survival advantage of married people.

Finally, our third contribution is to exploit the relatively long time span of the HRS to examine the time evolution of the socioeconomic gradient of expected longevity. We find relatively large increases in this gradient, although the precision of our estimates is low.

HRS Data

The Health and Retirement Study is a biennial panel of individual-level data collected from 1992 to 2010. The first wave of interviews included respondents born between 1931 and 1941, as well as their spouses regardless of age. New cohorts, both younger and older, have been introduced over the years. The original cohort is named HRS. The new cohorts are the AHEAD (introduced in 1993; respondents born before 1924), CODA (introduced in 1998; respondents born between 1924 and 1930), WB (introduced in 1998; respondents born between 1942 and 1947), and EBB (introduced in 2004; respondents born between 1948 and 1953). The overall sample contains respondents aged 51 or older, plus spouses of any age.

Individuals aged 50 in the HRS data set are from one of three cohorts: 1941–1942 (HRS cohort, observed in 1992), 1947–1948 (WB cohort, observed in 1998), and 1953–1954 (EBB cohort, observed in 2004). Figure 1 provides a complete description of the age-cohort structure of the HRS target population. The length of the arrow reflects the maximum age range in which we observe the individuals of a given year of birth. The dashed diagonal lines indicate the year of observation. The patent unbalanced entry of new cohorts generates an age structure of the target population that is very different in every sample year, which prevents a clear linking of our findings to a specific year.

Sample Selection

Our sample excludes individuals for whom we cannot obtain race, sex, or education, and individual-year observations for which we cannot obtain self-rated health or the survival status to the next interview. We keep individual-year observations with a positive sampling weight—which represent the civilian, noninstitutionalized U.S. population—and an age range of 50 to 94. We create separate samples for males and females because the slope of the socioeconomic gradient has been shown to be different for men and women (see, e.g., Elo and Preston 1996; Rogers et al. 2010). Our study focuses on white individuals because of the much smaller sample sizes for other racial/ethnic groups. Socioeconomic gradients of health exist in all groups, but they are of different magnitude; see Crimmins et al. (2004) and references therein.

We estimate survival rates and transition functions conditional on education, marital status, labor market status, wealth, income, smoking status and self-rated health. Some

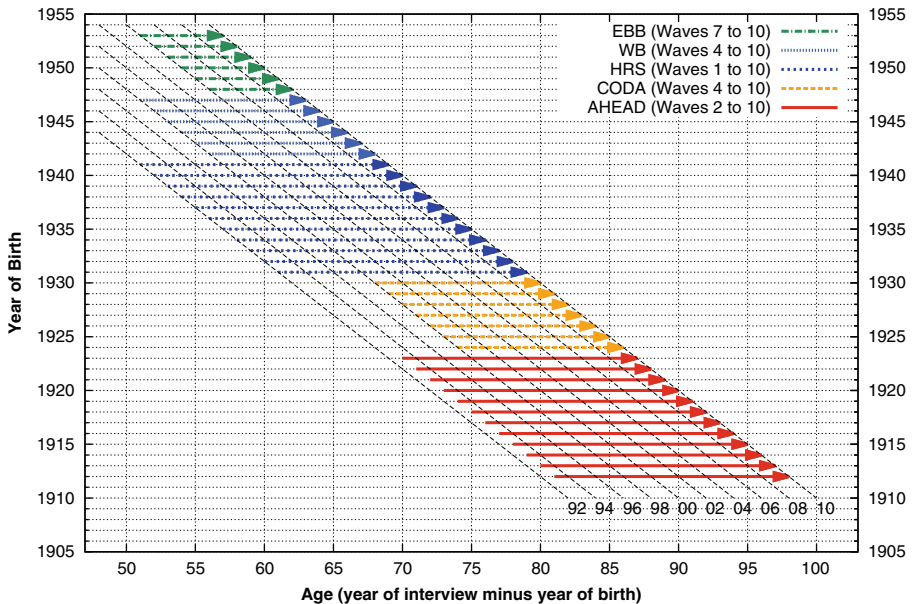


Fig. 1 Age structure of the HRS-eligible individuals. The arrows represent the maximum age range in which eligible individuals of a given year of birth are interviewed. Different HRS samples are denoted in the key. Age on the x-axis is year of interview minus year of birth; actual age may be one year younger

of these variables present missing data for a few observations. We do not omit observations with missing data other than education and self-rated health, so sample sizes in different estimations may differ slightly. We omit the observations with missing data for self-rated health because we want the samples used for the exercises with and without information about health to be identical. Overall, we have 53,362 individual-year observations for males and 67,453 for females, which correspond to 9,542 males and 11,236 females, respectively. Our sample period is 1992 to 2008 because no transition of any type can be observed in 2010. In Appendix A, we describe the variable definitions and provide more details about the sample selection.

Period Life Expectancies

The measurement of expected years of life for a subgroup of individuals of a given cohort is not an easy task. The crudest way of doing so is by aggregating the age-specific mortality rates—also known as life tables—of the population subgroup into *period life expectancies*. The period life expectancy at age 50 measures the average age of death for a hypothetical group of 50-year-olds, born at the same time and subject throughout their lifetime to the age-specific death rates of a particular time period, usually a given calendar year. The National Vital Statistical System (NVSS) computes the life tables for the U.S. population and reports period life expectancies for gender-race subgroups. The period life expectancy differs from expected longevity in that the latter accounts for possible changes in a person’s type and mortality rates over time. As such, the period life expectancy calculations provide a useful benchmark from which to assess the importance of these possible changes on the social gradient, with and without controls on self-reported health.

Due to the relatively small sample size of the HRS, we cannot use individuals born in different years to compute the death rate at every different age. Instead, we pool the data of all years to compute period life expectancies, using data from individuals in several cohorts to compute each age-specific survival probability. Because the HRS sample period is 1992 to 2008, we view our period life expectancies as a weighted average of those reported by the NVSS between these years.

Average Life Expectancy

We start by computing the life expectancy at age 50, using pooled data from the HRS to estimate separate age-specific two-year survival probabilities γ_a for white males and white females. These probabilities are estimated with a logit model that includes a linear term in age. Details on this and all the remaining estimations can be found in Appendix B. Let x_a be the number of people alive at age a out of a given initial population at age 50. Then, the life expectancy at age 50, e_{50} , can be computed as follows:

$$e_{50} = \sum_{a \in A} [a(1 - \gamma_a)x_a] + 1$$

$$x_{a+2} = \gamma_a x_a \quad \forall a \geq 50$$

$$x_{50} = 1.$$

Because the HRS is a biennial panel, all of our estimates refer to two-year periods. Due to the scarcity of data for very old individuals, we restrict our estimates to people up to age 94. Hence, we define $A \equiv \{50, 52, \dots, 94\}$. Note that our formula for life expectancy is unconventional. In actuarial sciences, life expectancy is typically defined as $e_{50} = \sum_{a \in A} a a \gamma_{50} + 1$, where $a \gamma_{50}$ is the probability of survival to age a for an individual aged 50 and is computed by use of the conditional survival probabilities γ_a . We prefer to keep our formulation to preserve comparability with the formulas of expected longevity used in the later sections of the article.

In the first column of Table 1, we report the point estimate of the average life expectancy and its standard error.² We find that life expectancy at age 50 is 78.8 years for white males and 82.9 years for white females. These numbers square well with the life expectancies computed with the life tables reported by the NVSS for 1992 to 2008. In particular, the NVSS life expectancies between 1992 and 2008 range from 77.0 to 79.3 for white males and from 81.7 to 82.9 for white females.³

The Socioeconomic Gradient in Life Expectancy

The age-specific mortality rates vary substantially with variables related to socioeconomic status. As we discussed in the Introduction, it is well known that the more educated, the wealth-rich, the income-rich, and the married have lower mortality rates. In addition, being active in the labor market has been shown to be related to lower mortality rates (see, e.g., Lin et al. 2003). We can aggregate these differences in mortality rates by computing life expectancies for each group. In particular, we compute life expectancies conditional on a characteristic $z \in Z \equiv \{z_1, z_2, \dots, z_n\}$ and obtain the difference $e_{50}(z_1) - e_{50}(z_n)$, where z_1 and the z_n are the most and the least advantaged types, respectively. We consider different sets Z of socioeconomic characteristics: education (college graduates, high school graduates, less than high school diploma), wealth (quintiles of the distribution of total household net worth per adult), nonfinancial income (quintiles of the distribution of labor income plus employer pensions and all government and social security transfers), labor market status (attached, which we define as working full-time or being unemployed searching for a job; semi-attached, which we define as working part-time or being semi-retired; and inactive, which includes inactive, retired, and disabled individuals), and marital status (married and its complement).⁴ In addition, we also consider smoking behavior, which is not a socioeconomic characteristic but is a risky behavior. However, because it is a habit, smoking tends to be very persistent over time; therefore, longevity predictions conditional on smoking status can be computed with the same methodology that we develop here. Finally, to illustrate the versatility of our methods, we also look at a four-category variable created by combining marital status and smoking. The interpretation of all these life expectancies is the expected age of death of a hypothetical group of

² The standard errors are obtained by drawing 25,000 samples of parameter values from the estimated asymptotic distribution of the model parameters and computing a life expectancy with each of them. See Appendix B for details.

³ The life expectancies we compute in the HRS should be somewhat larger than the ones reported by the NVSS because the HRS refers to the noninstitutionalized population. For instance, Brown et al. (2012) find that life expectancies at 65 in the HRS are about one year larger than in the NVSS.

⁴ See Appendix A for the exact definition of all these variables.

Table 1 Period life expectancies at age 50

	Life Expectancy Gradient							
	All	Education	Wealth	Nonfinancial Income	Labor Market Status	Marital Status	Smoking	Marital Status and Smoking
White Males	78.8 (0.2)	6.3 (0.7)	10.7 (0.8)	6.1 (0.9)	9.4 (1.0)	4.6 (0.6)	7.2 (0.5)	10.2 (0.9)
White Females	82.9 (0.2)	5.8 (0.7)	8.5 (0.6)	3.6 (0.7)	5.0 (1.8)	2.4 (0.4)	6.2 (0.5)	7.5 (0.7)

Notes: The first column reports the period life expectancy at age 50 for white males and white females. The remaining columns report the difference in life expectancy between the most and the least advantaged types for education, wealth, nonfinancial income, labor market status, marital status, smoking, and the combination of smoking and marital status. See Appendix A for the exact variable definitions. Standard errors are shown in parentheses.

50-year-olds with some characteristic $z = z_j$ who are subject throughout their lifetime to the z_j age-specific death rates of the current population alive.

To compute z -specific life expectancies, we first estimate age-specific two-year survival probabilities $\gamma_a(z)$ for every $z \in Z$. This involves estimating logistic regressions of survival against age, dummy variables for each $z \in Z$, and interaction terms between age and z in order to allow for the fall in survival due to age being different for different types z . The life expectancy $e_{50}(z_j)$ at age 50 for individuals whose z was equal to z_j at age 50 is then given by

$$e_{50}(z_j) = \sum_{a \in A} [a[1 - \gamma_a(z_j)]x_a] + 1$$

$$x_{a+2} = \gamma_a x_a \quad \forall a \geq 50$$

$$x_{50} = 1.$$

In Table 1, columns 2–7, we report the life expectancy differences at age 50 as well as the standard errors. We find huge socioeconomic gradients of life expectancies, all of them steeper for males than for females. At age 50, the differences in life expectancy for males and females, respectively, for each socioeconomic indicator are as follows: college graduates versus individuals without a high school diploma, 6.3 years for males and 5.8 years for females; for individuals at the top quintile versus the bottom quintile of the wealth distribution, 10.7 and 8.5 years; for individuals at the top versus bottom quintiles of the nonfinancial income distribution, 6.1 and 3.6 years; for individuals strongly attached to the labor force versus inactive individuals, 9.4 and 5.0 years; for married versus nonmarried individuals, 4.6 and 2.4 years; and for nonsmokers versus smokers, 7.2 and 6.2 years. Combining marital status and smoking also reveals large differences: the difference between married nonsmokers and nonmarried smokers is 10.2 years for males and 7.5 years for females. These results should not necessarily be

interpreted in terms of causality, given that selection of healthier individuals into better socioeconomic groups is known to matter.

Expected Longevities

Period life expectancy for individuals with a certain socioeconomic characteristic z can be a biased measure of the expected length of life of those individuals when the characteristic of interest z changes over time. For instance, the period life expectancy calculation for the top income quintile is likely biased upward because the calculation does not take into account the fact that some individuals will fall ill, drop out of the top income quintile, and then die (probably earlier than those who remain in the top income quintile). This bias is not a problem with education, which is fixed at age 50. However, wealth, income, marital status, and smoking behavior change substantially over the years.⁵

Accordingly, we start developing a measure of expected longevity at age 50 conditional on a given characteristic $z \in Z$ at age 50 that allows for changes in the characteristic z over the life cycle. In the next section, we further extend this concept using information about health to account for the changing composition of the health of the population whose expected longevity we want to measure. Two elements are required for these calculations: age-dependent survival rates conditional on z , $\gamma_a(z)$, and age-dependent transition probabilities for state z , $p_a(z'|z)$. Of course, the latter is not needed for education. Let $x_a(z)$ be the fraction of people who are alive and of type z at age a out of a given population at age 50. Expected longevity $\ell_{50}(z_j)$ at age 50 conditional on $z = z_j$ at age 50 is then computed as follows:

$$\ell_{50}(z_j) = \sum_{a \in A} \left[a \sum_{z \in Z} [1 - \gamma_a(z)] x_a(z) \right] + 1$$

$$x_{a+2}(z') = \sum_{z \in Z} p_a(z'|z) \gamma_a(z) x_a(z) \quad \forall z' \in Z, \forall a \geq 50$$

$$x_{50}(z_j) = 1 \text{ and } x_{50}(z) = 0 \quad \forall z \neq z_j.$$

We use the same estimates of $\gamma_a(z)$ as in the previous section, and we estimate multivariate logistic regressions with the same regressors in order to compute the transition matrices $p_a(z'|z)$. The interpretation of these expected longevities is the expected age of death of a hypothetical group of 50-year-olds with some characteristic $z = z_j$ who are subject throughout their lifetime to the age-specific type- z transition rates and age-specific survival rates conditional on z of the current population alive.

⁵ To point to particular examples, only 47 % of white males with wealth in the top quintile of the distribution at age 50 were in the same quintile by age 65 (with most of the movers going to the second quintile), and 88 % of white males who were married at age 50 were also married by age 65. Even more important are the changes in labor market status, because people clearly drop from the labor force as they age.

The Socioeconomic Gradient in Expected Longevity

In rows (1) of Table 2, we report the differences in expected longevity for the same socioeconomic categories as in Table 1 except for education, given that education does not change after age 50. We find substantial heterogeneity in expected longevity at age 50, with all differences statistically different from zero. The largest differences in expected longevity are between married men who are nonsmokers and nonmarried men who are smokers (4.6 years), and between men in the top and bottom quintiles of the wealth distribution (3.6 years). The socioeconomic gradient is generally more important for men than for women, except for nonfinancial income, which is relatively small at 0.8 years for both men and women.

Heterogeneity based on socioeconomic characteristics is substantially lower when changes in these characteristics after age 50 are accounted for, as shown by the large disparity in results between Tables 1 and 2. In fact, the estimates assuming static socioeconomic characteristics (period life expectancy) are 2 to 7 times higher for men and 2 to 6 times higher for women than the estimates that account for changes in socioeconomic characteristics (expected longevity). Because life expectancy is a

Table 2 Average expected longevity gradients at age 50 according to different measures of socioeconomic status

	Wealth	Nonfinancial Income	Labor Market Status	Marital Status	Smoking	Marital Status and Smoking
White Males						
(1) All individuals	3.6 (0.4)	0.8 (0.2)	2.0 (0.2)	2.2 (0.3)	2.2 (0.2)	4.6 (0.6)
(2) College graduates	2.8 (0.6)	0.7 (0.2)	1.5 (0.2)	1.6 (0.3)	1.6 (0.2)	3.5 (0.5)
(3) No high school diploma	3.7 (0.5)	0.8 (0.2)	2.3 (0.3)	2.9 (0.4)	2.7 (0.3)	5.6 (0.7)
(4) Education and z	8.4 (0.9)	6.5 (0.7)	7.9 (0.8)	8.9 (0.9)	8.0 (0.8)	10.8 (1.1)
White Females						
(1) All individuals	2.9 (0.2)	0.8 (0.1)	0.8 (0.1)	1.1 (0.2)	1.8 (0.3)	2.7 (0.3)
(2) College graduates	1.5 (0.3)	0.4 (0.1)	0.4 (0.1)	0.7 (0.1)	0.7 (0.2)	1.3 (0.2)
(3) No high school diploma	2.9 (0.3)	0.5 (0.1)	1.0 (0.2)	1.3 (0.3)	2.4 (0.3)	3.5 (0.5)
(4) Education and z	7.2 (0.7)	6.0 (0.7)	6.3 (0.7)	6.9 (0.7)	7.5 (0.7)	8.3 (0.8)

Notes: Rows (4) report the difference between individuals with a college degree and the most advantaged type z and individuals without a high school diploma and of the least advantaged type z . Standard errors are shown in parentheses.

measure of expected longevity that imposes an identity matrix for the transition $p_a(z'|z)$ of characteristic z , the lower the mobility between groups, the smaller the difference between the gradients in life expectancy and expected longevity. Our findings using the HRS data indicate the great empirical importance of accounting for mobility between groups.

The Socioeconomic Gradient Within Education Groups

More educated people tend to have more favorable socioeconomic characteristics (e.g., be richer and earn higher income, divorce less). In order to get a better sense of the extent to which the differences in expected longevity associated with different socioeconomic variables are different from each other, we also compute our measures of expected longevity by education group. In rows (2) and (3) of Table 2 we report these differentials for college graduates and for individuals without a high school diploma. We find that the differentials remain large within education groups, still more so for men than for women. In all cases, the differentials are larger within individuals without a high school diploma than within college graduates.

Finally, the longevity differences between college graduates with the most advantaged type z and individuals without a high school diploma with the least advantaged type are reported in rows (4) of Table 2. The differences are larger than the average difference by education group, underscoring the importance of characteristic z beyond education in estimating expected longevity.

Time Trends

The data from NVSS show an upward trend in life expectancies, with life expectancies at age 50 having increased by 2.2 years for white males and by 1.0 years for white females between 1992 and 2008. There is no reason to expect that all socioeconomic groups have shared equally in this improvement. Indeed, an increase in the educational gradient of mortality and life expectancy has been widely documented for the United States over these years (see, e.g., Meara et al. 2008; Montez et al. 2011; Preston and Elo 1995). To uncover possible time changes in the socioeconomic gradient of expected longevity, we add the calendar year to our estimates of the age-specific survival rates and the age-specific transition probabilities for types z . By adding a linear year term independent of age but z -type dependent, we allow for both survival probabilities and mobility between types to change over time, and to do so differently for different types. Instead, we restrict the time changes in survival and mobility to be homogeneous across ages.⁶ With these estimates, we can compute expected longevity as in Table 2 but specific to every year in our sample. The expected longevity of a particular year are consistent with individuals facing throughout their remaining life the mortality rates and the transition matrices of the given year.

⁶ Allowing for interactions between age, type, and year would increase the parameterization of our logit and multilogit models beyond tractability. In addition, the rationale for interacting time effects with age comes from the evidence that long-run gains in survival rates are different at different ages. However, these findings relate to both age differences and time intervals much wider than ours. See Lee and Carter (1992) for details.

In Table 3 we report the average life expectancy and the socioeconomic gradient in expected longevity for the first and last years of our sample period, as well as their changes over these 18 years. Our estimates show an increase of 2.1 years in the life expectancy at age 50 for white males between 1992 and 2008, with a standard error of 0.7. The corresponding increase reported by the NVSS is 2.2 years. Hence, our HRS sample captures the population trend for white males very well. By contrast, the life expectancy for white females falls by 0.3 years in our sample, whereas it increases by 1.0 years in the NVSS data.⁷

The estimated longevity differences have all been increasing over time with the exception of nonfinancial income. In particular, the educational difference increased by 1.7 years for white males and by 2.4 for white females, although the standard errors are of the same order of magnitude. This is consistent with Preston and Elo (1995), who found evidence of an increase in the education gradient of mortality rates between 1960 and the early 1980s.⁸ Montez et al. (2011), using the National Health Interview Survey Linked Mortality File, also found an increase in the mortality gradient for both white men and white women during the period 1986–2006. Meara et al. (2008) reported that the education difference in life expectancy at age 25 for white males and white females, respectively, increased by 0.9 years 1.1 years between the 1980s and the 1990s, and by 1.6 years and 1.9 between 1990 and 2000.⁹ This consistency in evidence supports our view about the value of looking at time trends with the HRS, despite the lack of precision of estimates using these data.

Our results for the other measures of socioeconomic status are novel, and hence they paint a wider picture. During the decades of the 1990s and 2000s, the college premium in the labor market and income inequality have increased (see Heathcote et al. 2010), and so has wealth inequality (see Díaz-Giménez et al. 1997, 2011). A tempting conclusion is that the increase in income and wealth inequality is behind the increase in the socioeconomic gradient of expected longevity. However, our results show that the gradients for marital status and smoking have also increased over this period. This increase might be due to the correlation between marital status or smoking with income-related variables, but it may also be due to an increase in the selection of long-lived individuals into marriage and nonsmoking.

Summary

Our findings to this point can be summarized as follows. First, socioeconomic differences in period life expectancies are very poor predictors of differences in expected longevities when the socioeconomic characteristics of interest change over the life cycle. Second, although education is the most important variable to predict longevity

⁷ Despite the fact that the standard error associated to the change in female life expectancy is large (0.6 years), this discrepancy between the HRS and the NVSS is worrisome. In a sense, we are stretching the HRS to its limits. As shown in Fig. 1, information on deaths for old individuals contain limited time variation. For instance, individuals aged 85 and older come only from the original AHEAD cohort. This problem is more acute for women, who die an average of four years later.

⁸ Preston and Elo (1995) showed that the education gradient of mortality rates computed with the National Longitudinal Mortality Study (NLMS) between 1979 and 1985 is larger than the one obtained by Kitagawa and Hauser (1973) with the death certificates and census data of 1960.

⁹ The results for the 1980s and 1990s are based on data from the NLMS, whereas the comparison between 1990 and 2000 is based on data from the death certificates in the Multiple Cause of Death files.

Table 3 Time trends in expected longevity at age 50

	Expected Longevity Gradient							
	All	Education	Wealth	Nonfinancial Income	Labor Market Status	Marital Status	Smoking	Marital Status and Smoking
White Males								
1992	77.6	5.1	2.4	0.9	1.6	1.5	1.6	3.6
	(0.3)	(1.4)	(0.4)	(0.2)	(0.2)	(0.4)	(0.3)	(0.8)
2008	79.6	6.8	4.5	0.8	2.3	2.7	2.7	5.4
	(0.5)	(1.1)	(0.7)	(0.2)	(0.3)	(0.5)	(0.4)	(0.8)
Δ	+2.1	+1.7	+2.1	-0.1	+0.7	+1.1	+1.1	+1.8
	(0.7)	(2.1)	(0.8)	(0.2)	(0.3)	(0.6)	(0.5)	(1.2)
1992 (NVSS)	77.1							
2008 (NVSS)	79.3							
Δ (NVSS)	+2.2							
White Females								
1992	83.0	4.6	2.5	0.8	0.4	0.6	1.1	1.4
	(0.4)	(1.3)	(0.3)	(0.1)	(0.2)	(0.4)	(0.3)	(0.5)
2008	82.8	7.0	3.1	0.7	1.1	1.5	2.4	3.9
	(0.3)	(1.0)	(0.3)	(0.1)	(0.2)	(0.3)	(0.5)	(0.6)
Δ	-0.3	+2.4	+0.6	-0.2	+0.7	+0.9	+1.3	+2.6
	(0.6)	(2.0)	(0.4)	(0.2)	(0.2)	(0.5)	(0.6)	(0.9)
1992 (NVSS)	81.9							
2008 (NVSS)	82.9							
Δ (NVSS)	+1.0							

Notes: The first column reports the period life expectancy at age 50 for different years. The remaining columns report the average expected longevity differences according to different measures of socioeconomic status for different years. Δ refers to the difference between 1992 and 2008. The first three rows in each panel refer to our own calculations with the HRS; the next three lines refer to the data reported by the NVSS. Standard errors are shown in parentheses.

differences, other socioeconomic variables—such as wealth, labor market status, and marital status—carry independent relevant information. Third, heterogeneity is greater within males than within females and within less-educated than within more-educated individuals. Fourth, these differences in expected longevity between socioeconomic groups increased substantially during the period we study.

Expected Longevity With Information About Health

We use self-assessed health to improve our measures of expected longevity. Self-assessed health has been found to be a very important determinant of survival probabilities even after socioeconomic characteristics and measured health conditions are controlled for (see, e.g., Idler and Benyamini 1997, 1999). In addition, self-rated health is an interesting measure of health because it is also present in several other data sets of

individual survey data commonly used by social scientists, such as the English Longitudinal Study of Ageing (ELSA).¹⁰

The HRS asks respondents to evaluate their general health level as excellent, very good, good, fair, or poor. We use these data to estimate age-dependent health h and type- z survival rates, as well as age-dependent joint health h and type- z transition functions. Due to sample size restrictions, we use our pooled data of all cohorts to compute these estimates. However, we compute expected longevity at age 50 only for those cohorts whose health distribution is observed at age 50—that is, for the cohorts 1941–1942, 1947–1948, and 1953–1954. Hence, these expected longevity assume that, conditional on health, the future survival rates by type and the transition functions of these three cohorts will be the same as those observed at the end of our study period for older cohorts. Because the underlying health distribution may differ, however, the actual predicted survival rates—and the transition functions—will also differ.

Average Expected Longevity

We start by computing the average expected longevity ℓ_{50}^h using information on health and without conditioning on any type variable z . To do so, we estimate age-dependent survival as a function of health $\gamma_a(h)$, an age-dependent health transition function $p_a(h'|h)$, and the initial health distribution $\varphi_{50}(h)$. Logistic and multinomial logistic regressions are estimated for the survival and transition functions, using as regressors a linear term in age, dummy variables for each $h \in H$, and interaction terms between age and h . We then compute the expected longevity ℓ_{50}^h as follows:

$$\ell_{50}^h = \sum_{a \in A} \left[a \sum_{h \in H} [1 - \gamma_a(h)] x_a(h) \right] + 1,$$

$$x_{a+2}(h') = \sum_{h \in A} p_a(h'|h) \gamma_a(h) x_a(h), \quad \forall h' \in H, \forall a \geq 50,$$

$$x_{50}(h) = \varphi_{50}(h).$$

This new measure of expected longevity, ℓ_{50}^h , is the expected remaining life of a given cohort of individuals that face the same age-dependent mortality rates conditional on health $\gamma_a(h)$ and the same age-dependent evolution of health $p_a(h'|h)$ as the current old, but may differ on the initial distribution of health $\varphi_{50}(h)$. Compared with the period life expectancy e_{50} computed earlier, the measure ℓ_{50}^h takes into account the possibility that the 50-year-olds born between 1941 and 1954 may face in the future mortality rates γ_a that are different from those faced by the current old, who were born earlier. The measure ℓ_{50}^h does so through the observed differences in health status, instead of relying on extrapolation from time series regressions as in the Lee and Carter (1992) type of methods. Although this approach does not attempt to extrapolate $\gamma_a(h)$ and $p_a(h'|h)$ to

¹⁰ Others include the Medical Expenditure Survey (MEPS), the National Health Interview Survey (NHIS), the National Health and Nutrition Examination Survey (NHANES), the National Longitudinal Study of Youth (NLSY), the Survey of Health Ageing and Retirement in Europe (SHARE), and the Panel Study of Income Dynamics (PSID).

future dates, it incorporates some of the improvements in survival over time: using our method, if the health distribution at age 50 of the 1941–1954 cohorts is better than it was for the older ones when they were aged 50, it must be the case that $\ell_{50}^h > e_{50}$.

In the first column of Table 4, we report the expected longevity ℓ_{50}^h . We find the expected longevity at age 50 to be 78.8 years for white males and 83.0 years for white females. These values are nearly identical to the actual life expectancies, e_{50} (see Table 1, column 1), which were 78.8 and 82.9, respectively. This indicates the absence of relevant differences in initial health at age 50 in favor of the 1941–1953 cohorts. Therefore, our findings suggest that if the large gains in life expectancy over the last years were to extend into the future, it would not be through the better health of the 50-year-olds, but rather through improvements over time of the age-dependent survival function ($\gamma_a(h)$) and the age-dependent transition function ($p_a(h|h)$).

The Socioeconomic Gradient in Expected Longevity

To use information on health to improve on our measures of the socioeconomic gradient of expected longevity, we need three elements: (1) the health distribution at age 50 for every type z , $\varphi_{50}(h|z)$; (2) the age-dependent joint health and characteristic z transition matrix, $p_a(z', h|z, h)$; and (3) the age-dependent survival rates conditional on health and characteristic z , $\gamma_a(z, h)$. We use the same logit and multilogit models as in the previous section, with dummy variables for each element in Z and H , and their interaction with age.¹¹ For the transitions, the outcome variable is given by all the elements in the set $Z \times H$, so our estimates allow for health changes to have a causal impact on socioeconomic characteristics. This is important because the potential impact of health on wealth, income, and labor market status has been largely documented (see, e.g., Currie and Madrian 1999; Smith 1999). In addition, Hu and Goldman (1990) found evidence of the importance of selection of less healthy individuals into single and divorced status. Let $x_a(z, h)$ be the fraction of people who are alive and of type z with health h at age a from a given population at age 50. Given these objects, we can build expected longevity $\ell_{50}^h(z_j)$ at age 50 conditional on $z = z_j$ as follows:

$$\ell_{50}^h(z_j) = \sum_{a \in A} \left[a \sum_{h \in H, z \in Z} [1 - \gamma_a(z, h)] x_a(z, h) \right] + 1,$$

$$x_{a+2}(z', h') = \sum_{h \in H, z \in Z} p_a(h', z' | h, z) \gamma_a(z, h) x_a(z, h) \quad \forall z' \in Z, \forall h' \in H, \forall a \geq 50,$$

$$x_{50}(z_j, h) = \varphi_{50}(h | z_j) \text{ and } x_{50}(z, h) = 0 \quad \forall z \neq z_j.$$

¹¹ Some authors choose to estimate the health and survival functions together through an ordered logit, thinking of death as an extra (and absorbing) health state (see, e.g., Yogo 2009). Our specification has two advantages. First, it is designed to estimate not only the effects of the type variables z into health but also the evolution of the type variables z and how this is affected by health itself. Second, it imposes less structure than an ordered logit model by allowing the marginal effect of any variable z on future health to differ from its marginal effect on mortality. This distinction is important. For instance, the effect of education on mortality is null after health is controlled for, but it is still an important determinant of the law of motion of health (see Appendix B). The decompositions in the next section are based precisely on this distinction.

Table 4 Expected longevities at age 50 with health status

	Expected Longevity Gradient							
	All	Education	Wealth	Nonfinancial Income	Labor Market Status	Marital Status	Smoking	Marital Status and Smoking
White Males								
All type-specific	78.8 (0.2)	6.6 (0.8)	4.3 (0.6)	1.5 (0.4)	3.8 (0.4)	2.6 (0.6)	2.9 (2.4)	5.8 (2.7)
(a) Only initial health		1.7 (0.2)	1.4 (0.2)	0.8 (0.2)	2.3 (0.3)	0.5 (0.2)	0.6 (0.1)	1.1 (0.3)
(b) Only transition		5.0 (0.4)	2.0 (0.4)	0.4 (0.2)	0.5 (0.1)	0.8 (0.5)	1.0 (1.9)	1.8 (2.6)
(c) Only mortality		0.0 (0.6)	1.0 (0.3)	0.3 (0.2)	0.7 (0.2)	1.2 (0.3)	1.2 (0.2)	2.7 (0.4)
White Females								
All type-specific	83.0 (0.2)	5.8 (0.5)	3.7 (0.4)	1.3 (0.2)	1.4 (0.2)	1.4 (0.5)	2.2 (0.3)	3.2 (1.4)
(a) Only initial health		1.1 (0.0)	1.1 (0.1)	0.7 (0.1)	0.8 (0.1)	0.3 (0.1)	0.2 (0.1)	0.4 (0.1)
(b) Only transition		4.7 (0.3)	1.9 (0.2)	0.5 (0.1)	0.3 (0.1)	0.8 (0.4)	1.0 (0.1)	1.6 (1.2)
(c) Only mortality		0.3 (0.5)	0.7 (0.2)	0.2 (0.1)	0.3 (0.1)	0.3 (0.1)	0.9 (0.2)	1.1 (0.2)

Note: The first column reports the expected longevity at age 50 for white males and white females, making use of the background information on health. The remaining columns report the difference in expected longevities between the most and the least advantaged types, making use of the background information on health. Rows (a), (b), and (c) measure the contribution of differences in health at age 50, differences in the evolution of health after age 50, and differences in mortality rates across socioeconomic groups for the overall gradient. Standard errors are shown in parentheses.

The statistic $\ell_{50}^h(z_j)$ shall be interpreted as the expected remaining life of a given cohort of individuals with characteristic $z = z_j$ at age 50 that face the same age-dependent mortality rates $\gamma_a(z, h)$ conditional on type z and health h , and the same age-dependent joint evolution of type z and health $p_a(h', z'|h, z)$ as the current old, but it may differ in the initial conditional distribution of health $\varphi_{50}(h|z_j)$. Therefore, the socioeconomic gradient of expected longevities computed with use of the health information, $\ell_{50}^h(z_1) - \ell_{50}^h(z_n)$, differs from the one computed before, $\ell_{50}(z_1) - \ell_{50}(z_n)$, by allowing the type- z mortality rates $\gamma_a(z)$ and the law of motion for z , $p_a(z'|z)$ to be different in the future. As discussed previously, it does so through the observed differences in the health distribution by type z at age 50 across cohorts.

In columns 2–8 of Table 4, we report the differences in expected longevities between individuals with different socioeconomic characteristics. In all cases, the differentials computed taking into account the information on health are larger than those computed without it (see Table 2 for comparison). For instance, the expected longevity difference for males due to education is 6.6

years when using the information on health versus only 6.3 years when not using it; for wealth it is 4.3 versus 3.6 years; for income, it is 1.5 versus 0.8 years; for labor market status, it is 3.8 versus 2.0 years; for marital status, it is 2.6 versus 2.2 years; and for smoking it is 2.9 versus 2.2 years. This suggests the existence of significant differences in the distribution of health across cohorts: for the cohorts born 1941–1954, the gap in health at age 50 between the least and the most advantaged types was larger than the gap of older cohorts. This implies that the health condition among the least advantaged types of the current old is better than it will be in the future when the 1941–1954 cohorts age.

Summary

We use information on self-rated health to estimate transitions of the socioeconomic characteristics allowing for the joint dependence of socioeconomic characteristics and health, and to calculate age-specific mortality rates of different health and socioeconomic groups, providing a more accurate assessment of the role of those characteristics. We find no further improvements in average mortality rates in the near future. In this sense, using expected longevity delivers no gains relative to the use of period life expectancies. If there were to be any improvement in life expectancies in the coming years, it would need to come from reductions of mortality conditional on health. However, the use of health information implies larger estimates of the socioeconomic gradient of expected longevity. Thus, we also expect the differences in mortality rates across socioeconomic groups to increase in the coming years. Unfortunately, the stringent data requirements that are needed to use health information prevent us from calculating time trends and expected longevity differences across finer partitions of the population, such as joint consideration of education and socioeconomic characteristics.

Decomposing the Socioeconomic Gradient in Expected Longevity

We exploit information about health to decompose the expected longevity gradients into three elements: (1) differences in health already present at age 50; (2) changes in health that developed after age 50; and (3) differences in two-year mortality rates unrelated to measured health. To perform this decomposition, we build expected longevity in which only one of three elements is allowed to depend on h . That is, instead of the triplet $\{\varphi_{50}(h|z), p_a(z', h'|z, h), \gamma_a(h)\}$, we use only one element in turn combined with the other two elements of $\{\varphi_{50}(h), p_a(h'|h), \gamma_a(h)\}$. The results in Table 4 show that when we look at education, approximately one-fourth of the life expectancy difference is due to differences in health at age 50 for different education groups, $\varphi_{50}(h|z)$ (see row (a) for both males and females). That is, college graduates report better self-rated health than individuals without a high school diploma. Given the average evolution of health over the life course, $p_a(h'|h)$, and the average mortality rates by health types, $\gamma_a(h)$, this difference in the initial distribution of health generates by itself a difference in life expectancy of 1.7 years for males and 1.1 years for females. The education-specific health transition matrix, $p_a(z', h'|z, h)$, accounts for about three-quarters of the education gap (see row (b) for both males and females). That is, the fact

that self-rated health deteriorates less for highly educated individuals generates by itself a life expectancy gap of 5.0 years for males and 4.7 years for females. Finally, the effect of education-specific mortality rates $\gamma_o(z, h)$ is almost null: 0.0 years for males and 0.3 years for females, but with standard errors of 0.6 and 0.5, respectively (see row (c) for both males and females). Indeed, in the underlying logit regressions, the effect of education on mortality rates is not statistically different from zero after we control for health; see Appendix B for details. The decomposition for the life expectancy gaps by wealth generates a similar albeit less dramatic pattern in which initial health differences account for roughly one-third of the effect, the health-protecting nature of wealth accounts for about one-half, but mortality rates account for less than one-quarter. In the case of income, the initial health distribution accounts for about one-half of the advantage of the top quintile, whereas the rest is shared by the health-protecting nature of income and by mortality specific rates, with the share of the latter being smaller.

The decomposition results are substantially different for smoking and marital status in that mortality rates are smaller for married or nonsmokers even after we control for self-rated health. The same is true when we look at both types together. In particular, type-specific mortality rates account for between one-third and one-half of the life expectancy differential for smoking and marital status for both males and females.

Finally, for labor market status, the initial distribution of health is very important: it accounts for 2.3 of 3.8 years for males and 0.8 of 1.4 years for females. This is consistent with the evidence that early retirement/inactivity is very much linked to health status.

The difference between education, wealth, and income on the one hand, and marital status and smoking on the other point to the limits of socioeconomic privilege. Although financial resources may be health-protecting, they lower mortality very little conditional on health (and not at all in the case of education). We find that after health has declined through some terminal condition, financial resources help very little. However, we find some advantages in terms of mortality conditional on health related to being married and a nonsmoker.

Conclusions

We developed a new methodology to compute differences in the expected longevity of individuals that have different socioeconomic characteristics—education, wealth, non-financial income, labor market attachment, marital status, and smoking status—at age 50. Our measure addresses the two main problems associated with the use of life expectancies: (1) that people's characteristics evolve over time, and (2) that there are time trends in mortality. Our methodology borrows from the literature on duration analysis: we estimate a hazard model for survival with time-varying stochastic endogenous covariates and use it to compute expected durations.¹² Because the expected life length after age 50 is much longer than our window of observation, we overcome the right-censoring problem by using data for individuals from older cohorts and exploiting information on health h . An ideal alternative way to compute longevity differences would be by following a cohort of individuals over time until they die. In this manner,

¹² See Lancaster (1990) for an overview of duration analysis.

the right-censoring problem would be completely eliminated. The problem with this approach is that it requires data that are unavailable in the United States and most other countries.

We uncovered substantial heterogeneity in expected longevity. For instance, a man in the top wealth quintile lives 4.3 more years than a man in the lowest wealth quintile; the corresponding figure for income is 1.5 years. However, a naive application of the methodology based on period life expectancies that does not address these two problems yields much larger values: 10.7 years for a 50-year-old male in the top wealth quintile versus another in the bottom quintile, with a corresponding difference of 6.1 years for the top income quintile versus the bottom quintile. Our methodology clearly provides much better answers when we are interested in comparing the expected duration of the lives of particular groups of people. Furthermore, the methodology is applicable to various other countries that have produced data sets with information similar to that provided in our data set; examples include the ELSA in the United Kingdom or the SHARE in continental Europe. Thus, our approach allows comparisons of the socioeconomic gradient of longevity between different countries, which can nicely complement the scarce international evidence on the socioeconomic gradient of life expectancies (see Majer et al. 2011).

This substantial heterogeneity in expected life duration matters for three reasons. First, the socioeconomic gradient in expected longevity probably dwarfs the welfare implications of the income differences accruing to different socioeconomic groups. Second, the redistributive power of public policies that are paid out as life annuities—such as retirement pensions, public medical assistance, or long-term care—may be partly eroded by the longer life expectancies of richer individuals. For instance, Fuster et al. (2003) show that the life expectancy differences between education groups makes the social security system more beneficial to the highly educated, despite the strong redistributive component introduced into the system. Finally, financial products, such as life annuities, life insurance, and medical insurance, are intimately related to the expected length of life. The measurement of expected longevity of different population subgroups may help assess whether the pricing of these products is actuarially fair, which is itself important to understand the take-up of these products.

Decomposing the differences in longevity into a fraction that is due to differences in self-perceived health at age 50, changes in health after age 50 and differences in mortality among groups with different socioeconomic characteristics shows that by far and large the most important component is the advantage that various socioeconomic groups have in preserving health. A salient finding here is that, while education and wealth seem to have little predictive power for two-year survival rates once self-assessed health is known, marital status does help predict survival. This raises important questions for further research in trying to understand the survival advantage of married people.

We also document an increasing time trend of the differences among socioeconomic groups during the sample period 1992–2008. As it is well known, income and wealth inequality have also risen during this period. We do not know whether these two phenomena are connected, but it is certainly worth exploring. At the same time, our results show that the socioeconomic gradients of expected longevity are larger when we use information on self-assessed health to predict the future mortality rates and the

future transition matrices between types of the current young. This implies that the socioeconomic differences in mortality are likely to widen in the coming years.

Finally, it is important to highlight that our methodology also has drawbacks. First, the Markov assumption, both in the transitions and the mortality rates, is certainly restrictive and rules out effects from the distant past. However, adding more lags to the estimates also has problems. Every added lag would require longer spells in the panel dimension and would therefore diminish the effective sample size. In addition, because states are quite persistent, the identification of the effects of every lag would be difficult, and the precision of the estimates would suffer. Second, any measurement error in socioeconomic characteristics would generate artificially high mobility between states, making the difference between the socioeconomic gradients in life expectancy and expected longevity artificially large. Hence, our method should be applied to data sets in which socioeconomic status is measured with confidence. Third, the HRS and related data sets do not provide large samples, which generates relatively large standard errors, particularly when looking at characteristics with many states.

Acknowledgments We are grateful for comments from Pierre-Carl Michaud, Pedro Mira, and Carolyn Wilkins; from attendants to seminars at University of Pennsylvania, Carlos III, and a LAEF conference on Health and Macroeconomics; and from the editor and referees. Inés Berniell provided excellent research assistance. Ríos-Rull thanks the National Science Foundation for Grant SES-1156228. The views expressed herein are those of the authors and not necessarily those of the Federal Reserve Bank of Minneapolis, or the Federal Reserve System.

Appendix A: Data

We use version M of the RAND files of the HRS, which covers 10 waves from 1992 to 2010.

Variable Definitions

Education: Variable RAEDUC provides five educational categories: no high school diploma, high school dropout with GED tests, high school graduate, high school graduate with some college, and college graduate. We pool the second, third, and fourth categories for a wider high school graduate category.

Wealth: The HRS provides several measures of assets and liabilities. We define wealth as total household net worth per adult, excluding second residences and mortgages on second residences because these two variables are unavailable for the third wave of the whole sample and for the second wave of the AHEAD subsample. Hence, total wealth is the sum of HwASTC, HwACHCK, HwACD, HwABOND, HwAOTH, HwAHOUS, HwARLES, HwATRANS, HwABSNS, and HwAIRA, minus HwDEBT, Hw-MORT, and HwHMLN. We then divide the resulting figure by 2 if the individual is married. We deflate our resulting variable by the CPI. Finally, in order to have a discrete version of the wealth variable, we classify every individual-year observation by the quintile of that individual-year observation in the wealth distribution over all individual-year observations, including both white males and white females. Hence, the top quintile represents individuals with household wealth over \$207,450; the cutoff points for the second

through fourth quintiles are, respectively, \$95,497, \$44,677, and \$11,597. All figures are in 1992 dollars.

Income: We use information on non-financial income, which is measured at the individual level. This is the sum of *RwIERN* (labor earnings), *RwIUNWC* (unemployment benefits), *RwIPENA* (employer pensions and annuities), *RwISRET* (retirement income from social security), *RwISSDI* (disability income from social security), and *RwIGXFR* (other government transfers). We deflate our resulting variable by the CPI and compute the quintile for the individual-year observation in the income distribution over all individual-year observations, including both white males and white females. Hence, the top quintile represents individuals with nonfinancial income over \$25,804; the cutoff points for the second through the fourth quintiles are, respectively, \$14,096, \$8,418, and \$4,554. All figures are in 1992 dollars.

Labor market status: Variable *RwLBRF* provides seven categories for the relationship of the respondent with the labor market. We reduce it to three: attached, semi-attached, and inactive. In the first category, we include individuals who are either working full-time or unemployed and looking for a job; in the second category, we include people working part-time or semiretired; in the third, we include individuals who are retired, disabled, or out of the labor force.

Marital status: We use variable *RwMSTAT* to classify as married those who indicate that they are either married or partnered. We classify the remaining respondents (the separated, divorced, widowed, and never married) as nonmarried.

Smoking: Variable *RwSMOKEN* reports whether the respondent is currently a smoker.

Health: Variable *RwSHLT*, reports five categories for the respondent's self-reported general health status: excellent, very good, good, fair, and poor.

Alive: For every individual-year observation, we need to determine whether the individual survives into the next wave. Every wave contains the variable *RwIWSTAT*, which indicates whether the individual actually responded to the interview as well as the mortality status of the respondent. Code 1 indicates that the respondent actually responded to the interview, so he/she is alive. Code 4 indicates that the respondent dropped from the sample, but a follow-up on the individual verified that he/she was alive. These two cases are the ones we count as alive. Code 5 indicates that the individual did not survive to the current wave. Finally, code 7 indicates that the individual withdrew from the sample (because of either the sample design or sample attrition) and his/her survival is not known. We classify code 7 cases as "missing."

Sample Selection

We start with all individuals in the age range 50 to 94. For every individual-year observation we record the relevant information for the next wave and drop individual-year observations in 2010—because information for the next wave is not available—

and individual-year observations with zero sampling weight. This yields 12,219 males and 15,081 females. We omit individuals with missing information for race (17 males and 16 females), nonwhite individuals (2,364 males and 3,351 females), and individuals with missing information for education (2 males and 1 female), which leaves 9,836 males and 11,713 females. Every individual is observed in several waves, and every individual-year observation is useful to estimate survival probabilities and transitions of our covariates. Overall, we have 59,167 and 74,372 individual-year observations, respectively, for males and females. We next omit individual-year observations for which we do not know survival status into the next wave (596 and 748 individual-year observations for males and females, respectively); this happens for some observations of individuals who could not be followed upon withdrawing from the sample (code 7 in *RwIWSSTAT*). Next, we drop individual-year observations with missing information on health (5,209 for males and 6,171 for females, which correspond to 286 and 331 different individuals, respectively). What is left is our working sample of 9,542 males who provide 53,362 individual-year observations and 11,236 females who provide 67,453 individual-year observations. Of these individual-year observations, we have 3,431 deaths for males and 3,235 for females, with average death rates of 6.4 % and 4.8 %, respectively.

Appendix B: Estimation of the Underlying Logistic Regressions

Survival Probabilities

We approximate parametrically the survival probabilities γ_a as a function of age a only, and $\gamma_a(z)$ as a function of age and some type $z \in Z$. We run logistic regressions of survival as follows:

$$\text{Prob}\left(\text{alive}_{t+2} = 1 \mid a_t, z_t\right) = \frac{e^{f(a_t, z_t)}}{1 + e^{f(a_t, z_t)}}$$

and

$$f(a_t, z_t) = \alpha_0 + \alpha_1 a_t + \sum_{i=2}^I \alpha_{2i} D_{z_t=z_i} + \sum_{i=2}^I \alpha_{3i} (D_{z_t=z_i} \times a_t),$$

where $D_{z_t=z_i}$ is a dummy variable that takes value 1 if $z_t = z_i$, and zero otherwise; alive_{t+2} is a dummy variable that takes value 1 if the individual is alive in the next wave, and zero otherwise. In Table 5, we show the results of these regressions for white males. The results of other regressions are available upon request. The categories into the set Z are always sorted from the most to the least advantaged type. We also tried specifications that add quadratic or cubic terms on age. In these specifications, the quadratic term is significant and improves the fit slightly. However, it does not change the computed life expectancies. Later, when we keep adding variables to the regression and interact them with age, it helps to have a parsimonious specification. In

Table 5 Logistic regressions for survival: White males

	Education	Wealth	Labor Market Status	Marital Status	Smoking	Marital Status and Smoking
Constant	10.53** (31.28)	10.89** (30.91)	9.02** (17.62)	9.76** (59.56)	10.48** (65.12)	10.46** (55.31)
$D_{z_t = z_2}$	-1.075** (-2.78)	-0.25 (-0.51)	0.25 (0.31)	-1.57** (-5.22)	-1.95** (-5.50)	-1.41** (-3.17)
$D_{z_t = z_3}$	-1.81** (-4.33)	-1.13* (-2.43)	-1.27* (-2.35)			-1.04** (-2.35)
$D_{z_t = z_4}$		-1.48** (-3.24)				-3.37** (-6.18)
$D_{z_t = z_5}$		-2.88** (-6.47)				
Age	-0.11** (-23.72)	-0.11** (-23.94)	-0.08** (-9.68)	-0.10** (-45.12)	-0.11** (-51.96)	-0.11** (-43.00)
Age $\times D_{z_t = z_2}$	0.01 [†] (1.86)	-0.00 (-0.05)	-0.00 (-0.38)	0.01** (3.79)	0.02** (3.32)	0.01 (1.54)
Age $\times D_{z_t = z_3}$	0.02** (2.96)	0.01 (1.58)	0.00 (0.41)			0.01 [†] (1.84)
Age $\times D_{z_t = z_4}$		0.01 [†] (1.79)				0.03** (4.02)
Age $\times D_{z_t = z_5}$		0.02** (3.99)				
<i>N</i>	56,322	56,299	55,590	56,299	56,037	56,014

Notes: The omitted dummy variable corresponds to the most advantaged category ($z_t = z_1$), and the rest of the dummy variables are ordered toward the least advantaged categories. Values in parentheses are *t* statistics.

[†] $p < .10$; * $p < .05$; ** $p < .01$

Fig. 2 we plot the survival rates for white males against age. In panel a, we plot the survival rates with the age term obtained through age dummy variables, and then also a linear, a quadratic, and a cubic polynomial. The quadratic polynomial improves the linear one in that it better captures the decline in survival in the very last years. In panel b, we plot the survival for college-educated males and individuals without a high school diploma with the age term captured either through age dummy variables or through a linear term, in both cases interacted with education. We see that the linear term is enough to capture the shape of the age profile in both education cases. In panels c and d, we do the same for health, and again the linear term captures the different shapes of each health group quite well.

When examining the expected longevity differences in different years, we need to compute time-dependent age-specific survival probabilities $\gamma_{a,t}$ and $\gamma_{a,t}(z)$. To do so, we include a variable *t* for calendar year, as well as its

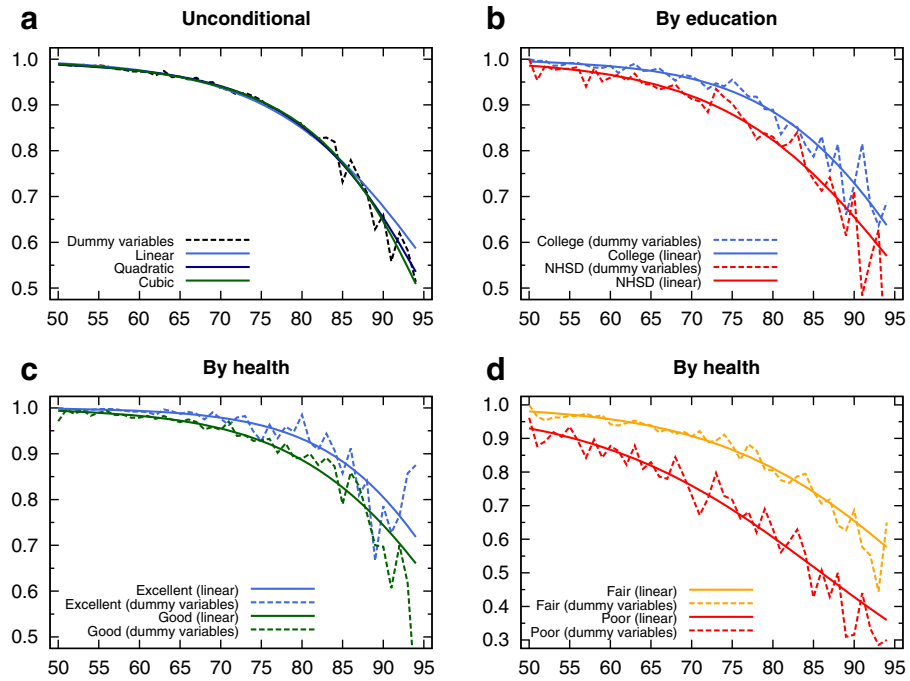


Fig. 2 Survival rates: parametric versus nonparametric age. Predicted yearly survival rates at a given age for the sample of white males. Because estimates correspond to two-year survivals, we report the square root of the predictions from our logit regressions. NHSD refers to no high school diploma

interaction with the type variable z . We do not, however, interact it with age. See footnote 6 for a discussion.

and

$$\text{Prob}\left(\text{alive}_{t+2} = 1 \mid t, a_t, z_t\right) = \frac{e^{f(t, a_t, z_t)}}{1 + e^{f(t, a_t, z_t)}}$$

$$f(t, a_t, z_t) = \alpha_0 + \alpha_1 a_t + \sum_{i=2}^I \alpha_{2i} D_{z_i=z_i} + \sum_{i=2}^I \alpha_{3i} (D_{z_i=z_i} \times a_t) + \alpha_4 t + \sum_{i=2}^I \alpha_{5i} (D_{z_i=z_i} \times t).$$

We do not report the results of these and the following regressions, but they are available upon request.

When examining the expected longevities by education group, we run the same survival regressions but for the given education subpopulation only.

To compute survival probabilities $\gamma_a(h)$ and $\gamma_a(h, z)$, we use the same logistic regression upgraded to include self-rated health, $h \in H$:

$$\text{Prob}\left(\text{alive}_{t+2} = 1 \mid a_t, h_t\right) = \alpha_0 + \alpha_1 a_t + \sum_{j=2}^J \alpha_{2j} (D_{h_t=h_j}) + \sum_{j=2}^J \alpha_{3j} (D_{h_t=h_j} \times a_t)$$

and

$$\begin{aligned} \text{Prob}\left(\text{alive}_{t+2} = 1 \mid a_t, z_t, h_t\right) &= \alpha_0 + \alpha_1 a_t + \sum_{j=2}^J \alpha_{2j} (D_{h_t=h_j}) \\ &+ \sum_{j=2}^J \alpha_{3j} (D_{h_t=h_j} \times a_t) + \sum_{i=2}^I \alpha_{4i} (D_{z_t=z_i}) + \sum_{i=2}^I \alpha_{5i} (D_{z_t=z_i} \times a_t). \end{aligned}$$

An important finding in this article is that, after we control for self-rated health, differences in educational attainment have very little predictive power for two-year-ahead mortality rates. To see this, we first look at the predictive power of each variable alone. In Fig. 3, we plot the predicted survival rates by health (panel a) and education (panel b) when the logistic regressions include only health or education variables. The clear result is that differences in self-rated health imply much larger differences in survival than do differences in education.

When we put both types of variables together in the same regression, we can use differences in the likelihood to test how much information education adds to health and how much information health adds to education. In Table 6 we report results from the three logistic regressions: only health variables, only education variables, and both together. The first row corresponds to survival depending only on education; the second row, to survival depending only on health; and the third row, to survival depending on both. As suggested by Fig. 3, the odds ratios are much larger when education categories are compared than when health categories are. Interestingly, when the two types of variables are added, the odds ratios of education become statistically not different from 1. Instead, the odds ratios for health become slightly larger when education is added to the regression. The results of the likelihood ratio test are clear. Compared with the model regression with both education and health, the constraint that all the coefficients on the health variables are zero is rejected strongly. Instead, the constraint that the education variables are zero is rejected at the 10 % confidence level but not at the 1 % level.

A visual inspection of these results comes from Fig. 4. In panel a, we reproduce the survival rates by education group as in Fig. 3, panel b. In panels b, c, and d of Fig. 4, we plot the predicted survival rates by education group within a health category. It is easy to see that within the health category, the role of education is minimal.

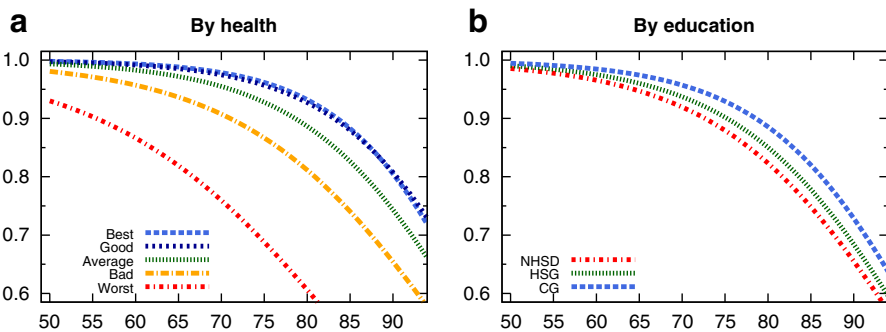


Fig. 3 Survival rates, by health and by education. Predicted yearly survival rates at a given age for the sample of white males. Because estimates correspond to two-year survivals, we report the square root of the predictions from our logit regressions. NHSD refers to no high school diploma; HSG, to high school graduates; and CG, to college graduates

Table 6 Logistic regressions for survival for white males: Health and education

	Odds Ratios				Likelihood Ratio Test	
	CG vs. NHSD		H_2 vs. H_4		χ^2	<i>p</i> Value
	Age 65	Age 75	Age 65	Age 75		
Only Education	4.46	4.30			1,897.89	.000
Only Health			171.19	170.90	9.32	.054
Both Together	1.08	1.16	202.29	202.02		

Notes: Each row corresponds to a logistic regression of survival against a linear term in age, the corresponding dummy variables (education dummy variables in the first row, health dummy variables in the second row, both sets of dummy variables in the third row), and the interaction of the linear term in age and the dummy variables. NHSD and CG refer to no high school diploma and college graduates respectively. H_2 and H_4 correspond to the second best and second worst health categories, respectively. The likelihood ratio tests in the first and second rows test the corresponding models against the model in the third row. Therefore, the null hypothesis in the first row is that all the health variables are zero and that only education matters (which is clearly rejected); and the null hypothesis in the second row is that all the education variables are zero and that only health matters (which is marginally rejected).

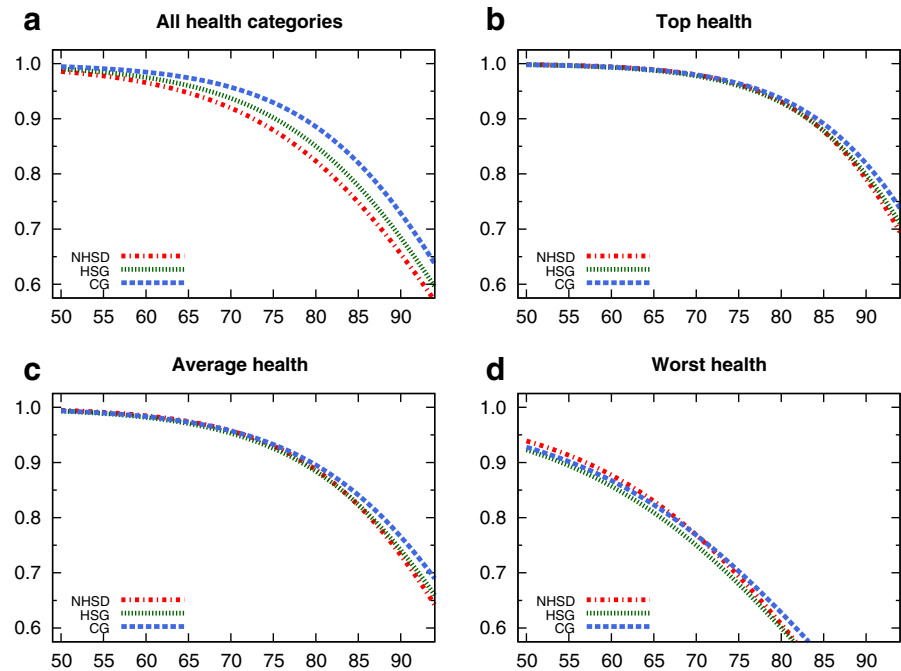


Fig. 4 Survival rates, by health and education jointly. Predicted yearly survival rates at a given age for the sample of white males. Because estimates correspond to two-year survivals, we report the square root of the predictions from our logit regressions. NHSD refers to no high school diploma; HSG, to high school graduates; and CG, to college graduates.

Transition Functions

We compute transition matrices $p_a(z'|z)$ by multivariate logistic regressions as follows:

$$\text{Prob}\left(z_{t+2} = z_1 \mid a_t, z_t\right) = \frac{1}{1 + \sum_{j=2}^I e^{f_j(a_t, z_t)}}$$

$$\text{Prob}\left(z_{t+2} = z_k \mid a_t, z_t\right) = \frac{e^{f_k(a_t, z_t)}}{1 + \sum_{j=2}^I e^{f_j(a_t, z_t)}} \quad \forall 1 < k < I,$$

with

$$f_k(a_t, z_t) = \alpha_{k0} + \alpha_{k1}a_t + \sum_{i=2}^I \alpha_{k2i}D_{z_t=z_i} + \sum_{i=2}^I \alpha_{k3i}(D_{z_t=z_i} \times a_t).$$

When we need to compute time-dependent transition matrices $p_{a,t}(z'|z)$, we add a variable for calendar year and interact it with the dummy variables for type:

$$\text{Prob}\left(z_{t+2} = z_1 \mid t, a_t, z_t\right) = \frac{1}{1 + \sum_{j=2}^I e^{f_j(t, a_t, z_t)}}$$

$$\text{Prob}\left(z_{t+2} = z_k \mid t, a_t, z_t\right) = \frac{e^{f_k(t, a_t, z_t)}}{1 + \sum_{j=2}^I e^{f_j(t, a_t, z_t)}} \quad \forall 1 < k \leq I,$$

with

$$f_k(t, a_t, z_t) = \alpha_{k0} + \alpha_{k1}a_t + \sum_{i=2}^I \alpha_{k2i}D_{z_t=z_i} + \sum_{i=2}^I \alpha_{k3i}(D_{z_t=z_i} \times a_t) + \alpha_{k4}t + \sum_{i=2}^I \alpha_{k5i}(D_{z_t=z_i} \times t).$$

Finally, when we need to compute transition matrices $p_a(h'|h)$ and $p_a(z', h'|z, h)$, we follow a similar approach. In the first case, we replace the $z \in Z$ by $h \in H$. In the second one, we create new dummy variables by combining $Z \times H$. This implies estimating very large models: the case for assets requires 25 outcome variables (5 asset categories multiplied by 5 health types). An alternative to using the transition matrices for the self-rated health would be to use an ordered logit. This approach is attractive because by imposing the structure of the ordered logit, we need to estimate much fewer parameters, and hence we

could potentially add more variables together. However, the restrictions imposed by the ordered logit are statistically rejected, so we use the multivariate logit.

Standard Errors

We compute the standard errors of the life expectancies and the expected longevities by simulation. Following Diermeier et al. (2003), we draw 25,000 vectors of parameters from the corresponding estimated asymptotic distribution and compute the desired life expectancies or expected longevities with each vector of parameters. This gives us a sample of 25,000 observations for each statistic of interest, and we report the mean and the standard deviation of this distribution. The vectors of parameters estimated in the logistic and multivariate logistic regressions are asymptotically normally distributed, with the mean given by the point estimates, and the variance-covariance matrix of the parameters also obtained from the estimation process. The initial distributions $\varphi_{50}(h)$ and $\varphi_{50}(h|z_j)$ follow multinomial distributions.

References

- Attanasio, O., & Hoynes, H. (2000). Differential mortality and wealth accumulation. *Journal of Human Resources*, 35, 1–29.
- Brown, D. C., Hayward, M. D., Montez, J. K., Hummer, R. A., Chiu, C.-T., & Hidajat, M. M. (2012). The significance of education for mortality compression in the United States. *Demography*, 49, 819–840.
- Brown, J. R. (2002). Differential mortality and the value of individual account retirement annuities. In M. Feldstein & J. B. Liebman (Eds.), *The distributional aspects of social security and social security reform* (pp. 401–446). Chicago, IL: University of Chicago Press.
- Crimmins, E. M., Hayward, M. D., & Seeman, T. (2004). Race/ethnicity, socioeconomic status and health. In N. B. Anderson, R. A. Bulatao, & B. Cohen (Eds.), *Critical perspectives on racial and ethnic differences in health in later life* (pp. 310–352). Washington, DC: National Academy Press.
- Currie, J., & Madrian, B. C. (1999). Health, health insurance and the labor market. In O. Ashenfelter & D. Card (Eds.), *Handbook of labor economics* (Vol. 3, pp. 3309–3416). Amsterdam, The Netherlands: Elsevier Science Publishers.
- Deaton, A., & Paxson, C. (1994). Mortality, education, income and inequality among American cohorts. In D. A. Wise (Ed.), *Themes in the economics of aging* (pp. 129–170). Chicago, IL: University of Chicago Press.
- Díaz-Giménez, J., Gover, A., & Ríos-Rull, J.-V. (2011). Facts on the distributions of earnings, income, and wealth in the United States: 2007 update. *Federal Reserve Bank of Minneapolis Quarterly Review*, 34(1), 2–31.
- Díaz-Giménez, J., Quadrini, V., & Ríos-Rull, J.-V. (1997). Dimensions of inequality: Facts on the U.S. distribution of earnings, income and wealth. *Federal Reserve Bank of Minneapolis Quarterly Review*, 21(2), 3–21.
- Diermeier, D., Eraslan, H., & Merlo, A. (2003). A structural model of government formation. *Econometrica*, 71, 27–70.
- Elo, I. T., & Preston, S. H. (1996). Educational differentials in mortality: United States, 1979–85. *Social Science & Medicine*, 42, 47–57.
- Fuster, L., Imrohoroğlu, A., & Imrohoroğlu, S. (2003). A welfare analysis of social security in a dynastic framework. *International Economic Review*, 44, 1247–1274.
- Heathcote, J., Perri, F., & Violante, G. (2010). Unequal we stand: An empirical analysis of economic inequality in the United States, 1967–2006. *Review of Economic Dynamics*, 1(13), 15–51.
- Hu, Y., & Goldman, N. (1990). Mortality differentials by marital status: An international comparison. *Demography*, 27, 233–250.
- Idler, E. L., & Benyamini, Y. (1997). Self-rated health and mortality: A review of twenty-seven community studies. *Journal of Health and Social Behavior*, 38, 21–37.

- Idler, E. L., & Benyamini, Y. (1999). Community studies reporting association between self-rated health and mortality. *Research on Aging*, 21, 392–401.
- Kitagawa, E. M., & Hauser, P. M. (1973). *Differential mortality in the United States: A study in socioeconomic epidemiology*. Cambridge, MA: Harvard University Press.
- Lancaster, T. (1990). *The econometric analysis of transition data*. Cambridge, UK: Cambridge University Press.
- Lee, R. D., & Carter, L. R. (1992). Modeling and forecasting U.S. mortality. *Journal of the American Statistical Association*, 87, 659–671.
- Lin, C. C., Rogot, E., Johnson, N. J., Sorlie, P. D., & Arias, E. (2003). A further study of life expectancy by socioeconomic factors in the National Longitudinal Mortality Study. *Ethnicity and Disease*, 13, 240–247.
- Majer, I. M., Nusselder, W. J., Mackenbach, J. P., & Kunst, A. E. (2011). Socioeconomic inequalities in life and health expectancies around official retirement age in 10 Western-European countries. *Journal of Epidemiology and Community Health*, 65, 972–979.
- Marmot, M. G., Shipley, M. J., & Rose, G. (1984). Inequalities in death-specific explanations of a general pattern? *Lancet*, 323, 1003–1006.
- Marmot, M. G., Smith, G. D., Stansfeld, S., Patel, C., North, F., Head, J., . . . Feeney, A. (1991). Health inequalities among British civil servants: The Whitehall II study. *Lancet*, 337, 1387–1393.
- Meara, E., Richards, S., & Cutler, D. (2008). The gap gets bigger: Changes in mortality and life expectancy, by education, 1981–2000. *Health Affairs*, 27, 350–360.
- Montez, J. K., Hummer, R. A., Hayward, M. D., Woo, H., & Rogers, R. G. (2011). Trends in the educational gradient of U.S. adult mortality from 1986 through 2006 by race, gender, and age group. *Research on Aging*, 33, 145–171.
- Preston, S. H., & Elo, I. T. (1995). Are educational differentials in adult mortality increasing in the United States? *Journal of Aging and Health*, 7, 476–496.
- Rogers, R. G., Everett, B. G., Saint-Onge, J. M., & Krueger, P. M. (2010). Social, behavioral, and biological factors, and sex differences in mortality. *Demography*, 47, 555–578.
- Singh, G. K., & Siahpush, M. (2006). Widening socioeconomic inequalities in us life expectancy, 1980–2000. *International Journal of Epidemiology*, 35, 969–979.
- Smith, J. P. (1999). Healthy bodies and thick wallets: The dual relationship between health and economic status. *Journal of Economic Perspectives*, 13, 145–166.
- Yogo, M. (2009). *Portfolio choice in retirement: Health risk and the demand for annuities, housing, and risky assets* (NBER Working Paper No. 15307). Cambridge, MA: National Bureau of Economic Research.