THE FUNCTIONAL FORM OF THE RELATIONSHIP BETWEEN EDUCATIONAL ATTAINMENT AND ALL-CAUSE ADULT MORTALITY IN THE UNITED STATES

Jennifer Karas Montez[1], Robert A. Hummer[1], and Mark D. Hayward[1]

[1] Population Research Center and Department of Sociology, University of Texas at Austin

* Acknowledgements: This paper was supported by a research grant from the Eunice Kennedy Shriver National Institute of Child Health and Human Development (1 R01-HD053696, PI Robert A. Hummer) and by infrastructure (5 R24 HD042849) and training (5 T32 HD007081) grants awarded to the Population Research Center at the University of Texas at Austin by the Eunice Kennedy Shriver National Institute of Health and Child Development. The authors thank John L. Boies and Norman J. Johnson from the U.S. Census Bureau for their technical assistance.
Please address any correspondence to: Jennifer Karas Montez, Population Research Center, University of Texas at Austin, 1 University Station G1800, Austin, TX (e-mail: jennkaras@prc.utexas.edu).

ABSTRACT

A vast literature has repeatedly documented the inverse association between education and U.S. adult mortality risks, but given little attention to identifying the optimal functional form of the association. A theoretical explanation of the association hinges on our ability to empirically describe it. Using the 1979-1998 National Longitudinal Mortality Study for non-Hispanic white and black adults aged 25-100 years during the mortality follow-up period (N=1,008,215), we evaluated 13 functional forms across race-gender-age subgroups to determine which form(s) best captured the association. Results revealed that a functional form that includes a linear decline in mortality risks from 0 to 11 years of education, followed by a step-change reduction in mortality risks upon attainment of a high school degree, at which point mortality risks resume a linear decline but with a steeper slope than that prior to a high school degree was generally preferred. The findings provide important clues for theoretical development of explanatory mechanisms: an explanation for the selected functional form may require integrating a credentialist perspective to explain the step-change reduction in mortality risks upon attainment of a high school degree, with a pure human capital perspective to explain the linear declines before and after that degree.

INTRODUCTION

Age-specific adult mortality rates in the United States declined in a very impressive fashion throughout the 20th century, leading to record highs in life expectancy in the first decade of the 21st century. At the same time, the latter half of the 20th century was also characterized by very well documented differences in adult mortality rates across levels of educational attainment (Elo and Preston 1996; Glied and Lleras-Muney 2008; Kitagawa and Hauser 1973; Lauderdale 2001; Lin et al. 2003; Molla, Madans, and Wagener 2004; Rogers, Hummer, and Nam 2000), a relationship that seems to be at least in part causal (Glied and Lleras-Muney 2008; Lleras-Muney 2005; Smith 2004). Thus, the relationship between educational attainment and adult mortality now attracts much greater concern and research attention than perhaps ever before. Moreover, it is clear that socioeconomic-related mortality differentials—including those by educational attainment—stand at the heart of the public health agenda of the United States. For example, the U.S. government has established that the reduction of educational differences in mortality is a key goal for *Healthy People 2010* and will likely be for *Healthy People 2020* when its goals, objectives, and action plans are released in January 2010 (U.S. Department of Health and Human Services 2000). Moreover, several recent studies point toward widening, rather than narrowing, educational gaps in U.S. adult mortality over the last two decades (Jemal et al. 2008; Meara, Richards, and Cutler 2008; Montez et al. under review).

While an enormous literature on the relationship between educational attainment and U.S. adult mortality has been developed over the last 20 years or so (for a review, see Hummer and Lariscy 2009), at the same time surprisingly little consideration has been devoted to what is the preeminent way of measuring educational attainment with regard to U.S. adult mortality. That is, what measurement scheme for educational attainment best captures the functional form of the relationship between educational attainment and the risk of adult mortality in the United States? In a related fashion, almost no attention has been given to how the functional form of the education-mortality relationship may or may not differ across demographic subgroups of the population. Yet it is well established that the education-mortality relationship varies in important ways across age groups and, to a lesser extent, by gender and race/ethnicity. Clearly, it is important to take this demographic heterogeneity into account in best documenting the functional form of the education-mortality relationship.

The overall goal of this paper, then, is to conduct a thorough examination of the functional form of the relationship between educational attainment and U.S. adult mortality. We focus on two questions: (1) Among a predetermined set of functional forms that is justified by previous research in this area, what form best describes the association between educational attainment and overall U.S. adult mortality? (2) Which one or more of these functional forms best describes the relationship between educational attainment and mortality for different age, gender, and race subgroups of the population? Answering these fundamental questions in a thorough fashion will help move the scientific and policy communities toward a richer understanding of one of the core relationships in social science today—that is, how group differences in the length of life within a population are structured by a principal component of socioeconomic status.

PREVIOUS RESEARCH

To date, just a few papers in the large education-mortality literature have given the specific and fundamental issue of the functional form of the relationship between educational attainment and mortality risk serious consideration. The work of Backlund et al. (1999), using an earlier version of U.S. data (from the 1980s) that we employ in this study, is the most thorough treatment of this topic to date. They specifically tested which of four different functional forms (one linear form and three discontinuous forms) best captured the relationship between educational attainment and mortality among working-aged (25-64) adults in the United States. Their results clearly found that a nonlinear relationship best depicted the relationship between educational attainment and mortality risk for U.S. individuals in this age range. More specifically, they found that for both women and men educational attainment was best specified in a rather simple trichotomous categorization (less than a high school diploma, a high school diploma but no college degree, or a college degree or more). For men, in particular, the reduced mortality risk associated with a college degree was especially pronounced. Note that their estimates were based on competing functional form models that controlled for the age, race, employment status, marital status, and household size of respondents. Such a strategy might be questioned based on the possible downstream influences of educational attainment on employment status, in particular, as well as on marital status. Thus, their estimates may not have been depicting the best gross, or overall, relationship between educational attainment and U.S. adult mortality given the control variables they included. Nonetheless, at least to date, it is the most complete and informative paper in this specific area of education-mortality research.

Other papers using U.S. data show that both continuous measures of educational attainment and categorical schemes yield valuable insights into the education-mortality relationship that are not easily apparent when only one or the other specification is used (Elo, Martikainen, and Smith 2006; Elo and Preston 1996; Zajacova and Hummer 2009). For example, Elo and Preston (1996) use a continuous specification of educational attainment to demonstrate that, on average, the log odds of mortality risk for U.S. adults drops roughly five percent for each additional year of education among individuals between the ages 25 of 64, and 2-3 percent for each additional year of education for persons aged 65 and over. Nonetheless, none of these papers specifically examined competing functional forms of the education-mortality relationship. It is clear, then, that a thorough examination of the functional form of the education-mortality relationship is needed in this research area, most specifically to best determine where in the educational distribution mortality risks are highest and lowest. Further, this is important to determine both among the population as a whole as well as among subgroups of the population defined by age, gender, and race/ethnicity. That is, the functional form of the education-mortality relationship may not be identical across of all of these subgroups because of the differing magnitudes by which educational attainment is associated with mortality risk across categories of age, gender, and race in the United States (Backlund et al. 1999; Crimmins 2005; Elo and Preston 1996; Kimbro et al. 2008; Lin et al. 2003; Zajacova and Hummer 2009).

CONCEPTUAL FRAMEWORK

Educational attainment is one of the principal components of socioeconomic status, along with occupation, income, and wealth. Nonetheless, there are clear reasons for using educational attainment as the key indicator of socioeconomic status when studying socioeconomic differentials in adult health and mortality (Hummer, Rogers, and Eberstein 1998; Preston and Taubman 1994). First, educational attainment is most often completed early in adult life and usually remains constant throughout the life course. In contrast, occupational status, income, and wealth accumulation vary in considerable ways throughout the life course and, at least in part, respond to health fluctuations (Smith 2004). Secondly, measures of educational attainment may be more relevant than other measures of socioeconomic status for individuals who have either retired from the workforce, are currently unemployed, or are out of the labor force. Third, survey respondents—who make up the individuals used in our analysis below—are more likely to report educational attainment (and with reasonable accuracy) in comparison to other socioeconomic indicators, particularly income and wealth. Finally, educational attainment typically precedes occupational status, income, and the accumulation of wealth in both the life course sense and causal sense (Mirowsky and Ross 2003).

Most studies focusing on education and adult mortality measure educational attainment using a single indicator of years of completed school. Such indicators are most often used in one of three distinct ways and, in most cases, are not evaluated against one another. First, educational attainment is sometime specified in a continuous fashion with values ranging from 0 to 17 or so (Zajacova 2006). Second, educational attainment is sometimes specified in a set of categories (e.g., 0-8, 9-12, 13+) that is partially based on important attainment thresholds, but that also plays to the strengths and weaknesses of official U.S. mortality data that are based on counts of death certificates in the numerator and Census estimates in the denominator (Christenson and Johnson 1995; Meara et al. 2008; Molla et al. 2004). Finally, education is also often specified in a set of categories that demarcate important cut-points (e.g., 0-11 years, 12 years, 13-15 years, 16+ years) in the educational distribution of degrees that are usually awarded after a certain number of years of attained education (Montez et al. 2009; Pappas et al. 1993; Phelan et al. 2004; Rogers et al. 2000).

Based on the work of Backlund and colleagues (1999), reviewed above, and a more broad-based review of the education-mortality literature (Hummer and Lariscy 2009), we identify here 13 potential functional forms that may best represent the relationship between educational attainment and U.S. adult mortality. Thus, this work moves beyond all current studies in this area by considering many potential specifications of educational attainment with regard to U.S. adult mortality These 13 forms are summarized in Table 1; our next set of subsections briefly describes each of these potential functional forms.

Table 1 about here

Model 1: Non-Parametric Model

The first of our 13 specifications is simply an unstructured, or non-parametric, model; this specification allows each measured level of education to vary in whatever way it might to best fit the data. That is, there are no groupings or constraints imposed on the measurement of education and each level of attainment may exhibit a higher or lower mortality risk in relation to a reference category (e.g., 12 years of education). This specification, then, is devoid of theoretical content and simply lets the data speak for itself. We evaluate the 12 more complex functional forms against this initial non-parametric specification.

Model 2: Linear Model

We refer to our second specification as a Linear Model. In this case, educational attainment is specified as a continuous variable, with the assumption being that each additional year of educational attainment brings with it an associated decrease in the log odds of mortality that is consistent throughout the education continuum (Zajacova 2006). This specification best reflects the idea that educational attainment is a pure form of human capital (Becker 1993; Ross and Mirowsky 1999); that is, each additional year allows individuals to better develop their cognitive functioning, increase their sense of control, improve their health behavior, acquire job-related skills, and develop the resources that are necessary to live healthier and longer lives. Such a specification implies that each additional year in the educational progression is no more or no less important than the previous one, at least in terms of mortality risk. Consequently, there are no particular gains associated with, e.g., completing high school or completing college.

Models 3-6: Step Change(s) with Zero Slopes

The next set of four models in Table 1 includes step-change specifications. In these models, mortality reductions are associated with increases in education, but only when such an increase propels individuals into an advanced educational category. That is, there are no mortality benefits of higher levels of education within specified educational categories. Several of these step-change models are credential-based (Collins 1979). For example, obtaining a high school degree or a college degree may increase the ability of individuals to qualify for certain jobs, earn greater income, and achieve a higher social status, all of which have life-long influences on health and age-specific risk of mortality. The findings of Backlund et al. (1999), discussed above, clearly fit a specification within this category. Recall that they found that the categories of less than a high school diploma (i.e., 0-11 years), a high school diploma but no college degree (i.e., 12-15 years), and a college degree or higher (i.e., 16+ years) best captured the relationship between educational attainment and mortality among U.S. working-aged adults during the 1980s. Related work on U.S. health outcomes, however, has found little support for a credential-based specification of educational attainment (Ross and Mirowsky 1999).

The specific distinctions in the four models within this general category are relatively modest and simply reflect differences in both the number of educational categories that are specified as well as exactly how these categories are comprised. Model 3 specifies two categories of educational attainment: everyone with a high school degree or less is included in the first category and everyone with anything more than a high school degree is included in the second category. This is most closely related to measurement schemes that are used with official

U.S. mortality data (Christenson and Johnson 1995; Molla et al. 2004). Model 4 includes three categories of educational attainment: less than a high school degree, a high school degree but nothing further, and anything more than a high school degree. Model 5 also includes three categories of attainment and is identical to the best specification identified by Backlund et al. (1999): less than a high school degree, a high school degree along with persons who have some college but no college degree, and persons with a college degree or higher. Finally, Model 6 includes four different categories of educational attainment: less than a high school degree, high school degree, some college but no college degree, and persons who have a college degree or higher.

Models 7-10: Step Change(s) with a Constant, Non-zero Slope

Our next set of four models contains hybrids of the step-change and linear approaches. In these models, each year of education is associated with a reduction in the log odds of mortality risk (as indicated in each of these models by a continuous variable of educational attainment) by the same amount, although an additional step-change decrease is experienced each time an important new level of attainment is acquired. Thus, step-changes in these models work above and beyond each additional year of education to influence mortality risk. As a result, this set of four models repeats the previous four categorical models (Models 3-6), but does so with the additional continuous variable of educational attainment included (Models 7-10). Backlund et al. (1999) specifically tested the third specification within this group of four models (our Model 9) but found that is fit less well than their strictly categorical specification (our Model 5).

Models 11-13: Step Change(s) with Varying Slopes

Our next set of three models also contains hybrids of the step-change and linear approaches. However, unlike Models 7-10, they allow the linear reduction in the log-odds of mortality risks associated with each additional year of education to be steeper within certain step-change demarcations than others. In other words, each year of education reduces mortality risks by the same amount until an important new level of education is attained, at which point mortality risks are further reduced by a step-change, and after which each year of education linearly reduces mortality risks by a different amount than the segment preceding the step-change. Models 11-13 reflect the idea that mortality risk reduction may occur at different paces along the education continuum if, for example, the primary mediators through which education reduces mortality risks vary along the way. For instance, each year of education prior to a high school degree may reduce mortality risks by X% through improved cognitive function, sense of control, and health behaviors; while each year of education beyond a high school degree may reduce mortality risks by Y% through more fulfilling and lucrative employment. To allow for varying slopes in between important attainment levels, these models include interaction terms between attainment levels and the continuous measure of education. Model 11 includes one slope from 0 years through a high school degree, a step-change when some college has been attained, and a second slope thereafter. Model 12 includes two step-changes and three slopes: it includes one slope from 0 to 11 years of education, a step-change at a high school degree followed by a second slope until a college degree is attained, at which point there is another step-change followed by a third slope from that point onwards. Backlund et al. (1999) tested this specification but found that is fit less well than their strictly categorical specification (our Model 5). Finally, Model 13 includes one slope from 0 through 11 years of education, and a step change at a high school degree after which a second slope is evident.

We now turn to our analysis. As mentioned previously, our analysis focuses on two questions: (1) Among the predetermined 13 functional forms, which form best describes the association between educational attainment and overall U.S. adult mortality?, and (2) Which one or more functional forms best describes the relationship between educational attainment and all-cause mortality for different age, gender, and race subgroups?

DATA AND METHODS

Data

We used the most recent version of the National Longitudinal Mortality Study (NLMS), which is a database well-suited for examining demographic and socioeconomic differentials in U.S. adult mortality (Rogot et al. 1992). The NLMS is created by linking adult respondents from the Current Population Surveys (CPS) to death records in the National Death Index (NDI). The CPS is a monthly survey of approximately 57,000 households that collects demographic and socioeconomic information from a nationally representative sample of the civilian non-institutionalized population of the United States (U.S. Census Bureau 2008). The NDI is a computerized database of all certified deaths in the United States since 1979. The most recent version of the NLMS links adult respondents from a 1980 census subsample and 23 waves of the CPS starting March 1979 and ending March 1998 to death records in the NDI through December 31, 2001. It contains roughly three million records and over 250,000 deaths. Our analyses are based on the private-use version of the 1979-1998 NLMS because, unlike the public-use version, it contains detailed information on date of birth, timing of interview, and date of death, which are important pieces of information for creating exact age and our person-year data structure.

The NLMS is the best dataset available for our present objective which requires: (1) a very large sample so that we can stratify our analyses by race-gender-age subgroups, and (2) detailed information on educational attainment throughout the education continuum. The NLMS provides an exceptionally large sample size and ample number of deaths for estimating race-gender-age stratified models. In addition, the NLMS provides information on the full range of educational attainment including, for example, single years of education below a high school degree on one extreme, and a range of graduate degrees including Masters Degrees and Doctorate Degrees on the other.

Sample

Our NLMS sample includes respondents from all 23 waves of the CPS, but excludes the 1980 census subsample whose education data is unavailable. We included non-Hispanic white and non-Hispanic black adults between 25 and 97 years of age at interview. We excluded groups other than non-Hispanic white and black adults because of the greater potential for education to be obtained abroad among groups with high levels of immigration, and because information on nativity is available for only a subset of CPS waves. We removed respondents 98 years of age at interview because the NLMS top-codes ages at 98 years. Roughly 3.5 percent of respondents did not provide their month and/or year of birth. For these respondents, we imputed month of birth by random assignment, and year of birth by subtracting their age from their year of interview. Among the resulting sample, we excluded 0.01% of adults who were missing information on educational attainment. These criteria resulted in a final analytic sample of 1,008,215 adults, with 164,289 (16.3%) of these individuals identified as subsequent decedents in the NDI.

Methods

As mentioned earlier, our objective was to systematically determine the best of 13 predefined functional forms of the relationship between education and all-cause mortality risk for each of 10 race-gender-age subgroups. We selected the 13 functional forms to represent the most likely forms within our overarching conceptual frameworks – linear, step changes with zero slopes, step changes with a constant non-zero slope, and step changes with varying slopes – as well as a pure nonparametric model. For example, we test the "step changes with zero slopes" framework using four specific functional forms. One form has a single step change after a high school degree; another form has a step change at a high school degree and then again after at least some college; another form has a step change at a high school degree and then again after a college degree; while the fourth form has step changes at a high school degree, at some college, and again at a college degree. The 13 functional forms are listed in Table 1.

Our education variable represents completed years of education. The NLMS provides a standardized measure of educational attainment across the 1979-1998 CPS years to account for changes in how the CPS recorded education before and after 1992. Prior to 1992, education was recorded in single years from 0 to 19+. Beginning in 1992, it was recorded in one to four year increments prior to 9th grade, one-year increments from 9th through 12th grade, and starting with high school the remainder of education levels were recorded as degrees obtained (e.g., associate degree, bachelor degree). From the NLMS standardized version, we created multiple measures of educational attainment. We created a continuous measure (x_{ed}) that includes 0, 2.5, 5.5, 7.5, 9, 10, 11, 12, 14, 16, and 19 years to estimate the linear functional form. Based on this continuous measure, we created 11 dichotomous variables (x_0, $x_{2.5}$, $x_{5.5}$,…x_{19}) to estimate the non-parametric functional form. We created several additional dichotomous variables to identify educational attainment subgroups for the remaining functional forms. These variables are listed below.

X_{lths} =1 for 0-11 years of education or 12 years but without a diploma or GED
X_{hs} =1 for a high school degree or GED
$X_{lths+hs}$ =1 for up to and including a high school degree or GED
X_{hs+sc} =1 for a high school degree or GED, or some college but no bachelor's degree
X_{sc} =1 for some college but no bachelor's degree
X_{co} =1 for a bachelor's degree or higher
X_{sc+co} =1 for some college but no bachelors degree, or a bachelor's degree or higher

For each of 10 race-gender-age subgroups, we estimated 13 logistic regression models to predict the annual odds of death from age and the specific functional form of education. We did not adjust for potential mediators such as income because our aim was to identify the best *gross* form of the relationship between education and U.S. adult mortality risks, and not the best form *net* of mediating pathways. The models are based on a person-year data structure in which we aged every adult by one year beginning with their interview year until their year of death, or the end of the follow-up period if they survived. In doing so, a small percentage of person-year records were aged beyond 100 years. To mitigate the chance that these records represented adults whose CPS records were difficult to match to the NDI, as opposed to true centenarians, we removed the 0.01% of person-year records for ages 100 years and older. With the exception of the linear functional form, all models removed the education group that contained a high school degree as the omitted reference. We did not adjust the models for the complex survey design of the CPS because previous research with the NLMS found that point estimates and standard errors are not materially affected (see Backlund et al. 1999), and because unweighted analyses are generally preferred when the weights are largely a function of the predictors (Winship and Radbill 1994). Finally, the 10 race-gender-age subgroups included two large groups (non-Hispanic white and black males 25+, non-Hispanic white and black females 25+) and eight subgroups defined by all combinations of two race/ethnicities (non-Hispanic white, non-Hispanic black), two genders (male, female), and two person-year age groups (25-64, 65+).

We determined the best of the 13 functional forms for the relationship between education and mortality risk for each race-gender-age subgroup using the Bayesian Information Criterion (BIC). We chose the BIC because it is preferable to a Likelihood Ratio χ^2 Test when the sample size is large, as well as when the models to be compared are non-nested (Raftery 1995). We identified the best functional form for each race-gender-age subgroup as the logistic regression model with the smallest BIC value. The BIC value is calculated from the following equation:

$$BIC = -1*[(-2LL_0 – (-2LL_1)] + (\text{number of non-intercept model parameters})*\ln(N)$$

where $–2LL_0$ reflects the deviance associated with the intercept-only model, $–2LL_1$ reflects the deviance associated with the specified model, and N reflects the sample size (Raftery 1995).

Access to the private-use version of the NLMS is highly restricted to protect the confidentiality of certain survey information. Thus, we did not have access to the individual-level NLMS data. Instead, we conducted our analyses in coordination with U.S. Census Bureau staff. This entailed jointly developing the SAS programs for the analysis, which the Census Bureau staff processed and subsequently provided to us in the form of SAS output files.

RESULTS

One of the main advantages of using the 1979-1998 NLMS for our objectives is the unusually large sample size available for race-gender-age stratified analyses, and this advantage is illustrated in Table 2. Table 2 shows the numbers of deaths and respondents from individual, not person-year, records. Even subgroups that tend to be under-represented in national surveys have a relatively large sample size in the NLMS. For instance, our analytic sample contains roughly 1000 non-Hispanic black women 65 years of age and older with more than a high school education, who experienced 350 deaths during the follow-up period.

Table 2 about here

Figure 1 displays the log-odds coefficients for individual years of education estimated from the non-parametric functional form (Model 1). Attainment of a high school degree and a bachelor's degree – potentially important attainment levels for mortality risk reduction– are indicated by enlarged markers. The figure illustrates a few expected trends. For example, the inverse association between education and mortality risks appears steeper among adults 25-64 years than among those 65 and older within each race-gender group. A large body of literature has documented this pattern and debated its causes such as increasing returns to education among younger cohorts, decreasing returns to education with age due to disengagement from social stratification systems and/or a greater influence of biological aging processes, compositional changes within educational strata, or simply an artifact of mortality selection. The figure also suggests a few other, perhaps unexpected, trends that will be formally evaluated in the following section. For instance, there appears to be a relatively linear decline in the log-odds coefficients from 0 to 11 years, followed by a step-change reduction at a high school degree, with some subgroups groups (particularly those 25-64 years) displaying a steeper linear decline from a high school degree onwards.

Figure 1 about here

We now formally examine which of the 13 functional forms best describes the association between education and overall U.S. adult mortality. The first two columns of Table 3 contain rankings of the 13 functional forms - from best to worst based on BIC values - for non-Hispanic white and black males 25 years and older in column 1, and for non-Hispanic white and black females 25 years and older in column 2. The results show that Model 13 has the smallest BIC value for both groups, and thus best describes the association between education and U.S. adult mortality risks among the 13 functional forms. The table also identifies models with the next two smallest BIC values in light shade. If these values are within two BIC units of the optimal form, they are considered to provide a similarly good fit to the data (Raftery 1995) and are identified in the table by dark shades like the optimal form. The pseudo R^2 for both models is a reasonably high 0.15. Thus, these results corroborate our visual inspection of Figure 1 by identifying the best of the 13 functional forms for U.S. adult women and men overall as a linear decline in mortality risks across 0 to 11 years of education, perhaps followed by a step-change at a high school degree, and a different linear decline from a high school degree onwards.

Table 3 about here

We now examine which one or more functional forms best describe the association between education and all-cause mortality for different race, gender, and age subgroups. We first discuss the results for white adults shown in columns 3 through 6 of Table 3. The best functional form identified for all white females and older white males is, again, Model 13. For younger white males, Model 13 was identified as the second best model. For these men, Model 9 performed somewhat better and includes a constant linear decline in mortality risks throughout the education continuum with additional step-change reductions at a high school degree and a college degree. A more general inspection of the rankings for white adults reveals other interesting patterns. First, the non-parametric and linear models perform poorly. The linear model has a ranking between 8 and 10 of the 13 models, and for older white adults, it actually performs worse than the non-parametric model. Second, the set of models described by "step-changes with zero slopes" also performs poorly. Third, all models that combine a high school degree with less than high school provide the worst fit to the data compared with other models in each respective set. For example, Model 3 performs worse than Models 4-6, Model 7 performs worse than Models 8-10, and Model 11 performs worse than Models 12-13. These results then reveal a fundamental difference in mortality risks before and after a high school degree for white adults. Taken together, the results for white adults clearly identify Model 13 as the preferred functional form.

The results for black adults are shown in columns 7 through 10 in Table 3. The pattern is not as consistent as it was for white adults, although there tends to be a clustering of better performing models within the "step-changes with zero slope" set of models. For older black adults, choosing an optimal functional form may be a moot point given that there is little mortality risk reduction with increasing education levels, which is illustrated in Figure 1 and supported by the small pseudo R^2 values (0.02) for the best functional forms in Table 3. For younger black men, Model 5, which includes step-changes at a high school degree and at a college degree but zero slopes throughout, performed best. For these men, Model 13 (which was the best form overall for white adults) performed second best. The results for young black women select a model that contains step-changes at high school, and again at some college, with zero slopes throughout as the best form (Model 4); and the same step-changes but with constant, non-zero slopes as the second best model (Model 8). Model 13 is ranked as the fourth best model. Similar to white adults, the linear model fits poorly (except for older black males) and the distinction between 0 to 11 years of education versus a high school degree is stark. Taken together, the results for young black adults identify models with step-changes and zero slopes as the best fitting models, with Model 13 emerging as a close alternative. The results for older black adults indicate that there may not be a true "optimal" form, given the weak association and generally poor fitting models as reflected by the low pseudo R^2 values.

Table 4 about here

Given that Model 13 is the optimal form for white adults, and it is a good form for black adults, we provide log-odds coefficients for this model by race-gender-age groups in Table 4. For the two overall population groups in columns 1 and 2, the coefficients confirm that the step-change in mortality risk reduction at a high school degree is statistically significant (-0.519 for males[1], –0.210 for females) and that the linear decline in mortality risk reduction is shallower across 0 to 11 years of education (-0.011 for males, -0.004 for females) than it is from a high school degree onwards (-0.073 for males, -0.043 for females). Columns 3 through 6 show similar findings for the four subgroups of white adults. One exception is that the step-change reduction at a high school degree was not statistically significant for white women 25-64 years, which is surprising given the visually impressive step-change in Figure 1. As expected, the coefficients for black adults in columns 7 –10 are not consistently significant due to the weaker inverse association between education and mortality risks among older black adults and because Model 13 was not the top ranked form for black adults. However, consistent with the findings for white adults, the linear decline in mortality risks among younger black adults is shallower across 0 to 11 years of education (-0.019 for men, -0.023 for women) than it is starting with a high school degree onwards (-0.074 for men, -0.103 for women), and the step-change reduction in mortality risks at a high school degree is significant for women (-0.567).

[1] This coefficient reflects a statistically significant step-change. However, the actual step-change in the log-odds between 11 years of education and a high school degree for men is –0.236, which can be calculated from [(-8.350 – 0.073x11 – 0.519 + 0.062x11) – (-8.350 – 0.073x12)]. The actual step-change between 11 and 12 years can be calculated likewise for all other subgroups.

DISCUSSION

Our systematic investigation of 13 predefined functional forms for the association between educational attainment and U.S. adult mortality risks revealed a preference for a form that includes a linear decline in mortality risks between 0 and 11 years of education, followed by a step-change reduction in mortality risks upon attainment of a high school degree, at which point risk reduction continues to decline linearly but with a steeper slope than that prior to a high school degree. This functional form best described the association when aggregating non-Hispanic white and black males 25 years and older, and for their female counterparts. Our more detailed race-gender-age stratified analyses found that the same form was also preferred for white men and women. This form also provided a good fit for black adults, ranking in the top two to five best forms. However, the best functional form for black adults appears to fall within the set that includes step-changes with zero slopes. Given that most analyses combine white and black adults, or stratify with the intent to compare results, our findings suggest that the functional form described above (Model 13 in our analysis) is generally preferred. That said, analyses that specifically focus on non-Hispanic black adults may want to explore the set of models described by step-changes with zero slopes to evaluate whether they generate more informative results.

The consistency of the findings across race-gender-age subgroups is striking. The fact that one form consistently emerged as the best (or near best in some cases) form across race, gender, and age subgroups indicates that the mechanisms through which education shapes mortality risks may be fairly universal, and consistent over time and age, among U.S. adults. For instance, despite the long-standing debate regarding the causes of seemingly smaller educational differentials in health and mortality risks among older compared with younger adults (e.g., House et al. 1994; Lauderdale 2001; Lynch 2003; Mirowsky and Ross 2008; Ross and Wu

1996), we found that the association between education and mortality risks has the same functional form for *both* age groups among white adults. Likewise, despite evidence that education reduces mortality risks marginally more for older white men than women in the United States (Montez et al. 2009), we found that the association between education and mortality risks has the same functional form for these men *and* women.

For several reasons, our findings depart from Backlund et al (1999) who conducted the most thorough treatment of this topic to date. They found that the association for both men and women was best specified with a "step-changes with zero slopes" categorization (less than a high school diploma, a high school diploma but no college degree, or a college degree or more). A likely explanation for the discrepant findings is that we examined the *gross* association between education and mortality risks, while Backlund and colleagues examined the association *net* of household size, employment status, and marital status. Because these factors are correlated with education, controlling for them may have had the unintended consequence of explaining away a portion of the association, leaving only a residual portion available for investigation. Nevertheless, our conclusions do not depart dramatically. Indeed, Backlund stated that the "step-changes with varying slopes" model was a statistically close alternative form for men. Furthermore, consistent with our findings, they found that the "step-changes with varying slopes" model provided a significant better fit than the linear model for both men and women.

The current study not only provides compelling evidence for one particular functional form, it also provides important clues for further theoretical development and testing of explanatory mechanisms. For instance, the theoretical explanation for the preferred functional form may require integrating a credentialist perspective (Collins 1979) to explain the step-change reduction in mortality risks upon attainment of a high school degree, along with a human capital perspective (Becker 1993; Mirowsky and Ross 2003) to explain the linear declines before and after high school attainment. However, the search for explanatory mechanisms will need to employ a dataset that contains a wide range of potential causal mechanisms such as income, occupation, wealth, health behaviors, social ties, and psychosocial resources. The NLMS is the best dataset available for identifying the optimal functional form across race-gender-age subgroups due to its large sample size and its rich data on educational attainment throughout the education continuum. However, it is not suited for investigating causal mechanisms that link education to the risk of death. For instance, income and occupation are collected once at the time of survey, even though mortality follow-up extends for up to 23 years. Measures of cumulative exposure across long-term income and occupation trajectories are needed to capture the lifetime consequences of these mediators on adult mortality risks (Deaton 1999; Moore and Hayward 1990). In addition, information on health behaviors is severely limited in the NLMS. Data on smoking behavior was collected in just 5 of the 23 waves. Data on other important health behaviors such as alcohol consumption and physical exercise, as well as psychosocial resources such as a sense of control and marital history are unavailable. Regardless of these limitations, the strengths of the NLMS for the present objectives clearly outweigh its weaknesses.

Future research on explanatory mechanisms should also evaluate specific causes of death. In this initial investigation, we selected all-cause mortality risks for both conceptual and practical reasons. All-cause mortality is one of the best indicators of overall population health, and thus seemed to be the logical choice for launching this work. However, the best functional form for all-cause mortality is not necessary the best form for cause-specific mortality due to distinct etiological processes. In conclusion, this study significantly advances our understanding of the functional form of the association between educational attainment and U.S. adult mortality risks, and by extension, expands our understanding of the association between mortality risks and socioeconomic status more generally. By capitalizing on the strengths of the NLMS for the present objectives, we have provided the requisite groundwork for which future research on explanatory mechanisms can build upon.

TABLE 1. Predetermined Functional Forms of the Association Between Education and All-Cause Mortality Risks

Non-Parametric Model

1. $\text{log-odds} = b_0 + b_1x_0 + b_2x_{2.5} + b_3x_{5.5} + \ldots + b_{11}x_{19}$

Linear Model

2. $\text{log-odds} = b_0 + b_1x_{ed}$

Step Changes with Zero Slopes

3. $\text{log-odds} = b_0 + b_1x_{lths+hs} + b_2x_{sc+co}$
4. $\text{log-odds} = b_0 + b_1x_{lths} + b_2x_{hs} + b_3x_{sc+co}$
5. $\text{log-odds} = b_0 + b_1x_{lths} + b_2x_{hs+sc} + b_3x_{co}$
6. $\text{log-odds} = b_0 + b_1x_{lths} + b_2x_{hs} + b_3x_{sc} + b_4x_{co}$

Hybrid Model A (Step changes with a constant, non-zero slope)

7. $\text{log-odds} = b_0 + b_1x_{ed} + b_2x_{lths+hs} + b_3x_{sc+co}$
8. $\text{log-odds} = b_0 + b_1x_{ed} + b_2x_{lths} + b_3x_{hs} + b_4x_{sc+co}$
9. $\text{log-odds} = b_0 + b_1x_{ed} + b_2x_{lths} + b_3x_{hs+sc} + b_4x_{co}$
10. $\text{log-odds} = b_0 + b_1x_{ed} + b_2x_{lths} + b_3x_{hs} + b_4x_{sc} + b_5x_{co}$

Hybrid Model B (Step changes with varying slopes)

11. $\text{log-odds} = b_0 + b_1x_{ed} + b_2x_{lths+hs} + b_3x_{sc+co} + b_4x_{ed}*x_{lths+hs} + b_5x_{ed}*x_{sc+co}$
12. $\text{log-odds} = b_0 + b_1x_{ed} + b_2x_{lths} + b_3x_{hs+sc} + b_4x_{co} + b_5x_{ed}*x_{lths} + b_6x_{ed}*x_{hs+sc} + b_7x_{ed}*x_{co}$
13. $\text{log-odds} = b_0 + b_1x_{ed} + b_2x_{lths} + b_3x_{ed}*x_{lths}$

Notes: See page 14 for a description of the education variables. All models control for age. Except for the linear model, all models exclude the variable that contains a high school degree as the omitted reference.

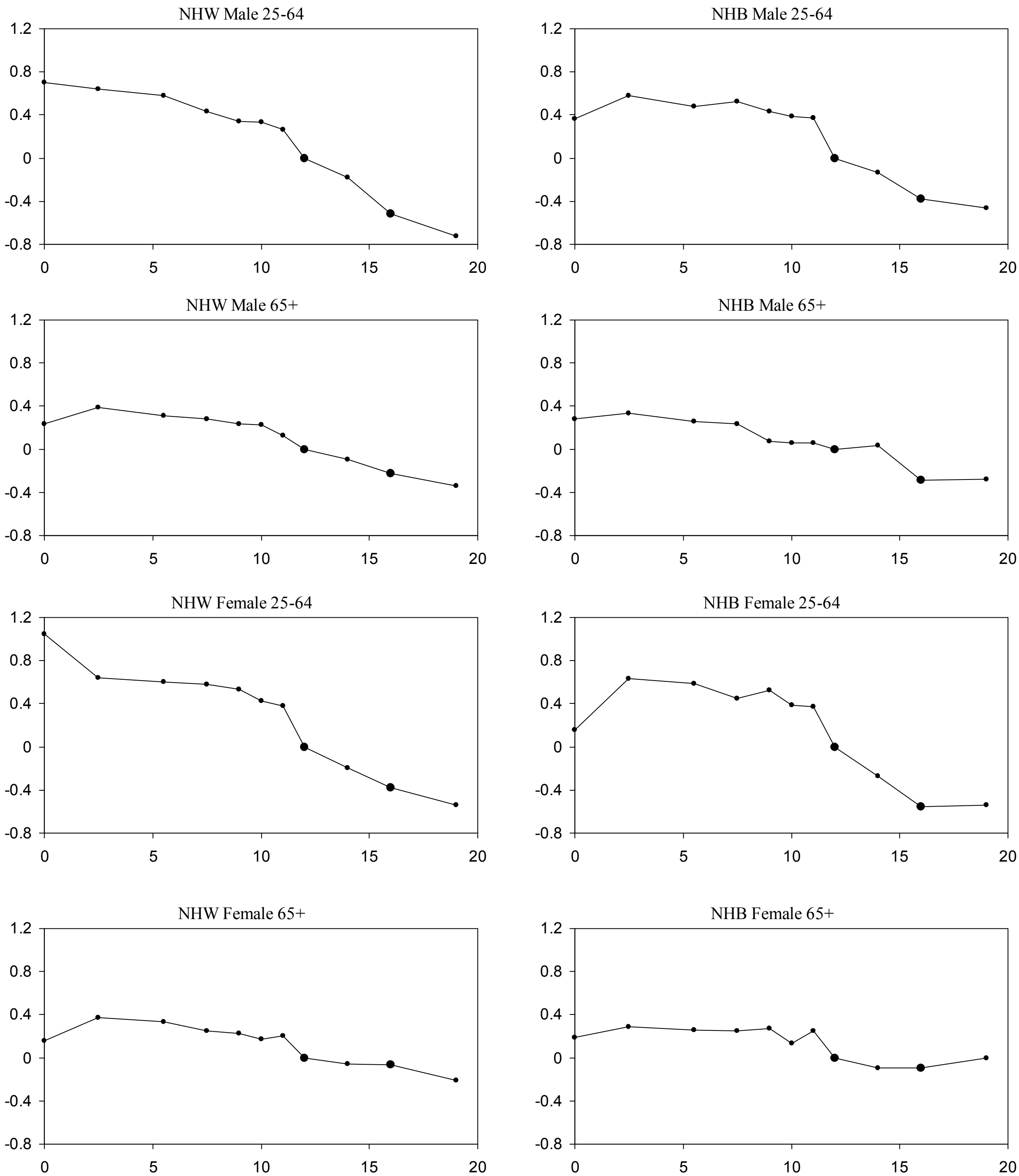
TABLE 2. Sample Sizes and Number of Deaths by Race-Gender-Age[1] Group and Education

	Non-Hispanic White Males				Non-Hispanic Black Males			
	25-64 years		65+ years		25-64 years		65+ years	
Education	Deaths	N	Deaths	N	Deaths	N	Deaths	N
Less than HS	11,761	56,873	23,220	35,894	2,717	11,162	2,847	4,763
HS	12,824	128,255	10,315	21,090	1,252	12,331	385	836
Some college	4,852	71,878	3,950	8,882	503	6,309	138	340
College	4,966	96,400	4,606	10,858	264	4,076	117	280
Total	34,403	353,406	42,091	76,724	4,736	33,878	3,487	6,219

	Non-Hispanic White Females				Non-Hispanic Black Females			
	25-64 years		65+ years		25-64 years		65+ years	
	Deaths	N	Deaths	N	Deaths	N	Deaths	N
Less than HS	7,861	56,529	25,378	47,548	2,418	14,620	3,230	6,787
HS	10,481	165,130	13,702	36,185	1,220	17,777	535	1,594
Some college	2,985	77,891	5,013	13,550	363	8,896	176	530
College	2,075	75,285	3,735	9,651	226	5,551	174	464
Total	23,402	374,835	47,828	106,934	4,227	46,844	4,115	9,375

[1]Age reflects age at interview. Sample sizes and deaths reflect individual respondents, not person-years.

FIGURE 1. Log-odds Coefficients for Nonparametric Levels of Education (Functional Form 1)



Note: High school and college degree are indicated with enlarged markers.

TABLE 3. Ranking of Functional Forms for the Association between Education and Mortality Risks within each Race-Gender-Age[1] Group based on the Bayesian Information Criterion (BIC)

	NHW/B male 25+	NHW/B female 25+	NHW male 25-64	NHW male 65+	NHW female 25-64	NHW female 65+	NHB male 25-64	NHB male 65+	NHB female 25-64	NHB female 65+
1. Nonparametric	4	9	10	6	11	7	12	13	12	13
2. Linear (ed)	12	12	8	10	9	10	9	1	10	7
Step changes with zero slopes										
3. lths+hs, sc+co	13	13	13	13	13	13	13	12	13	11
4. lths, hs, sc+co	10	6	12	12	7	6	3	10	1	1
5. lths, hs+sc, co	7	8	11	8	8	9	1	8	5	2
6. lths, hs, sc, co	5	2	9	7	6	8	5	11	3	4
Step changes with constant, nonzero slopes										
7. ed, lths+hs, sc+co	11	10	7	9	10	11	10	2	9	8
8. ed, lths, hs, sc+co	8	5	5	5	2	2	6	6	2	3
9. ed, lths, hs+sc, co	6	7	1	2	3	3	4	3	6	6
10. ed, lths, hs, sc, co	3	4	3	3	4	4	7	7	7	9
Step changes with varying slopes										
11. ed, lths+hs, sc+co, ed*lths+hs, ed*sc+co	9	11	6	11	12	12	11	4	11	10
12. ed, lths, hs+sc, co, ed*lths, ed*hs+sc, ed*co	2	3	4	4	5	5	8	9	8	12
13. ed, lths, ed*lths	1	1	2	1	1	1	2	5	4	5
Pseudo R^2 for optimal functional form[2]	0.15	0.15	0.06	0.04	0.06	0.05	0.05	0.02	0.05	0.02

[1] Age reflects person-year age

[2] Pseudo $R^2 = [-2LL_0 - (-2LL_1)]/(-2LL_1)$ where $-2LL_0$ and $-2LL_1$ are the deviances of the intercept and specified models, respectively.

Notes: For each subgroup, the optimal functional form is the form with the smallest BIC value (shown by darkest shade). The next two smallest BIC values are indicated by a light shade. If either BIC was within 2 points of the optimal form, they are indicated by the dark shade instead of the light shade, because a difference in BIC of less than two points is not practically meaningful (Raftery 1995).

TABLE 4. Log-odds Coefficients for Functional Form 13, "Step Change at High School with Varying Slopes," by Race-Gender-Age[1]

	NHW/B Male 25+	NHW/B Female 25+	NHW male 25-64	NHW male 65+	NHW female 25-64	NHW female 65+	NHB male 25-64	NHB male 65+	NHB female 25-64	NHB female 65+
Intercept	-8.350**	-9.543**	-8.266**	-7.934**	-9.536**	-9.606**	-7.457**	-6.106**	-7.739**	-7.603**
Age[1]	0.078**	0.082**	0.083**	0.070**	0.089**	0.081**	0.067**	0.045**	0.070**	0.053**
Non-Hispanic Black (reference is white)	0.143**	0.163**	---	---	---	---	---	---	---	---
Education (measured in years)	-0.073**	-0.043**	-0.110**	-0.051**	-0.085**	-0.025**	-0.074**	-0.042*	-0.103**	-0.009
Less than a High School Degree	-0.519**	-0.210**	-0.542**	-0.175**	-0.115	0.093†	-0.284	-0.142	-0.567†	0.198
Education*Less than a High School Degree	0.062**	0.039**	0.064**	0.029**	0.040**	0.006	0.055**	0.016	0.080**	0.003
Slope before high school degree	*-0.011*	*-0.004*	*-0.046*	*-0.022*	*-0.045*	*-0.019*	*-0.019*	*-0.026*	*-0.023*	*-0.006*
Slope starting with high school degree	*-0.073*	*-0.043*	*-0.110*	*-0.051*	*-0.085*	*-0.025*	*-0.074*	*-0.042*	*-0.103*	*-0.009*
Deaths	84,561	79,068	18,422	57,932	11,468	59,299	3,109	5,098	2,629	5,672
Person Years	6,519,840	7,669,169	4,504,231	1,474,417	4,804,487	2,077,311	425,657	115,535	600,261	187,110
Pseudo R^2 $[(-2LL_0 - (-2LL_1)]/(-2LL_0)$	0.15	0.15	0.06	0.04	0.06	0.05	0.05	0.02	0.05	0.02

†p<0.05, *p<0.01, **p<0.001

[1] Age reflects person-year age.

REFERENCES

Backlund, E., P.D. Sorlie, and N.J. Johnson. 1999. "A Comparison of the Relationships of Education and Income with Mortality: The National Longitudinal Mortality Study." *Social Science & Medicine* 49:1373-1384.

Becker, G.S. 1993. *Human Capital: A Theoretical and Empirical Analysis with Special Reference to Education* 3rd Edition. The University of Chicago Press.

Christenson, B.A. and N.E. Johnson. 1995. "Educational Inequality in Adult Mortality: An Assessment with Death Certificate Data from Michigan." *Demography* 32:215-229.

Collins, R. 1979. *The Credential Society: A Historical Sociology of Education and Stratification*. New York: Academic Press.

Crimmins, E.M. 2005. "Socioeconomic Differentials in Mortality and Health at the Older Ages." *Genus* 61:163-176.

Deaton, A. 1999. "Inequalities in Income and Inequalities in Health." *NBER Working Paper 7141. National Bureau of Economic Research*.

Elo, I., P. Martikainen, and K. Smith. 2006. "Socioeconomic Differentials in Mortality in Finland and the United States: The Role of Education and Income." *European Journal of Population* 22:179-203.

Elo, I.T. and S.H. Preston. 1996. "Educational Differentials in Mortality: United States, 1979-85." *Social Science & Medicine* 42:47-57.

Glied, S. and A. Lleras-Muney. 2008. "Technological Innovation and Inequality in Health." *Demography* 45:741-761.

House, J.S., J.M. Lepkowski, A.M. Kinney, R.P. Mero, R.C. Kessler, and A.R. Herzog. 1994. "The Social Stratification of Aging and Health." *Journal of Health and Social Behavior* 35:213-234.

Hummer, R.A. and J. Lariscy. 2009. "Educational Attainment and Adult Mortality." Pp. XX-YY forthcoming chapter in *International Handbook of Adult Mortality*, edited by R. G. Rogers and E. M. Crimmins. Springer.

Hummer, R.A., R.G. Rogers, and I.W. Eberstein. 1998. "Sociodemographic Differentials in Adult Mortality: A Review of Analytic Approaches." *Population and Development Review* 24:553-578.

Jemal, A., E. Ward, R.N. Anderson, T. Murray, and M.J. Thun. 2008. "Widening of Socioeconomic Inequalities in U.S. Death Rates, 1993-2001." *PLoS ONE 3(5):e2181* 3:1-8.

Kimbro, R.T., S. Bzostek, N. Goldman, and G. Rodriguez. 2008. "Race, Ethnicity, and the Education Gradient in Health." *Health Affairs* 27:361-372.

Kitagawa, E.M. and P.M. Hauser. 1973. *Differential Mortality in the United States: A Study in Socioeconomic Epidemiology*. Cambridge, MA: Harvard University Press.

Lauderdale, D.S. 2001. "Education and Survival: Birth Cohort, Period, and Age Effects." *Demography* 38:551-561.

Lin, C.C., E. Rogot, N.J. Johnson, P.D. Sorlie, and E. Arias. 2003. "A Further Study of Life Expectancy by Socioeconomic Factors in the National Longitudinal Mortality Study." *Ethnicity & Disease* 13:240-247.

Lleras-Muney, A. 2005. "The Relationship between Education and Adult Mortality in the United States." *Review of Economic Studies* 72:189-221.

Lynch, S.M. 2003. "Cohort and Life-Course Patterns in the Relationship between Education and Health: A Hierarchical Approach." *Demography* 40:309-331.

Meara, E.R., S. Richards, and D.M. Cutler. 2008. "The Gap Gets Bigger: Changes in Mortality and Life Expectancy, by Education, 1981-2000." *Health Affairs* 27:350-360.

Mirowsky, J. and C.E. Ross. 2003. *Education, Social Status, and Health*. New York: Aldine De Gruyter.

———. 2008. "Education and Self-Rated Health. Cumulative Advantage and Its Rising Importance." *Research on Aging* 30:93-122.

Molla, M.T., J.H. Madans, and D.K. Wagener. 2004. "Differentials in Adult Mortality and Activity Limitation by Years of Education in the United States at the End of the 1990s." *Population and Development Review* 30:625-646.

Montez, J.K., M.D. Hayward, D.C. Brown, and R.A. Hummer. 2009. "Why Is the Educational Gradient in Mortality Steeper for Men?" *Journal of Gerontology: Social Sciences* 64:625-634.

Montez, J.K., R.A. Hummer, M.D. Hayward, H. Woo, and R.G. Rogers. under review. "The Beginning of the End or Just the Beginning? Recent Trends in the Educational Gradient of U.S. Adult Mortality by Race, Gender, and Age Group."

Moore, D.E. and M.D. Hayward. 1990. "Occupational Careers and Mortality of Elderly Men." *Demography* 27:31-53.

Pappas, G., S. Queen, W. Hadden, and G. Fisher. 1993. "The Increasing Disparity in Mortality between Socioeconomic Groups in the United States, 1960 and 1986." *New England Journal of Medicine* 329:103-109.

Phelan, J.C., B.G. Link, A. Diez-Roux, I. Kawachi, and B. Levin. 2004. ""Fundamental Causes" Of Social Inequalities in Mortality: A Test of the Theory." *Journal of Health and Social Behavior* 45:265-285.

Preston, S.H. and P. Taubman. 1994. "Socioeconomic Differences in Adult Mortality and Health Status." Pp. 279-318 in *Demography of Aging*, edited by L. G. Martin and S. H. Preston. Washington DC: National Academy Press.

Raftery, A.E. 1995. "Bayesian Model Selection in Social Research." *Sociological Methodology* 25:111-163.

Rogers, R.G., R.A. Hummer, and C.B. Nam. 2000. *Living and Dying in the USA: Behavioral, Health, and Social Differentials of Adult Mortality*. New York: Academic Press.

Rogot, E., P.D. Sorlie, N.J. Johnson, and C. Schmitt. 1992. A Mortality Study of 1.3 Million Persons by Demographic, Social, and Economic Factors: 1979-1985 Follow-Up. Bethesda, MD: NIH Publication 92-3297.

Ross, C.E. and J. Mirowsky. 1999. "Refining the Association between Education and Health: The Effects of Quantity, Credential, and Selectivity." *Demography* 36:445-460.

Ross, C.E. and C.-L. Wu. 1996. "Education, Age, and the Cumulative Advantage in Health." *Journal of Health and Social Behavior* 37:104-120.

Smith, J.P. 2004. "Unraveling the Ses-Health Connection." *Population and Development Review* 30(Suppl):108-132.

U.S. Census Bureau. 2008. Current Population Survey Annual Social and Economic Supplement. Washington D.C. Accessed at http://www.census.gov/apsd/techdoc/cps/cpsmar08.pdf.

U.S. Department of Health and Human Services. 2000. *Healthy People 2010: Understanding and Improving Health*. Washington, D.C.: U.S. Government Printing Office.

Winship, C. and L. Radbill. 1994. "Sampling Weights and Regression Analysis." *Sociological Methods & Research* 23:230-257.

Zajacova, A. 2006. "Education, Gender, and Mortality: Does Schooling Have the Same Effect on Mortality for Men and Women in the U.S.?" *Social Science & Medicine* 63:2176-2190.

Zajacova, A. and R.A. Hummer. 2009. "Gender Differences in Education Effects on All-Cause Mortality for White and Black Adults in the United States." *Social Science & Medicine (forthcoming)*.