The Non-Linear Relationship Between Education and Mortality*

Bethany G. Everett, University of Colorado

David H. Rehkopf, University of California at San Francisco

Richard G. Rogers, University of Colorado

Robert Hummer, University of Texas, Austin

September 20, 2009

*Acknowledgements: Funded in part through the Eunice Shriver National Institute of Child Health and Human Development (NICHD) research grant #R01 053696. We thank the NICHD-funded University of Colorado Population Center (grant R21 HD51146) for administrative and computing support

ABSTRACT

Much research has found a strong relationship between education and mortality. Methodologically, this has been accomplished by using a linear term or a series of dummy variables to measure the relationship between education and mortality, despite the fact that there exists evidence that the relationship is not linear, and may vary between sexes and race/ethnic groups. We assess the relationship between education and mortality using Cox proportional hazard models with a penalized spline on education. We find a linear relationship only among non-Hispanic whites. Among non-Hispanic blacks and foreign born Mexican men and women we see decreased mortality risk at both the low ends and high ends of education, with a mortality risk peak occurring between 9 and 12 years of education. This results provide important insights into how education operates across race/ethnic groups and its relationship to mortality.

BACKGROUND

For almost four decades researchers have been examining the relationship between education and mortality. Beginning with Kitigawa and Hauser, a strong graded relationship between education and mortality has emerged (1973). The pathways through which education improves health are both multifaceted and complex. Education not only improves income and access to insurance and health care (Becker 1975; Elo and Preston 1996; Hayward and Gorman 2004; Rogers, Hummer, and Nam 2000), but also improves social networks, shapes lifestyle and behaviors, and also works through more abstract concepts like increased feelings of personal mastery and self-efficacy (Lantz et al., 1998; Link and Phelan, 1995; Mirowsky and Ross, 1998, 2003). Because of the variety of pathways through which education influences health, it has been described as the "fundamental cause" of health disparities in the United States.

Education also reduces stress. Persons with higher levels of education have both increased social networks through which to reach out to individuals for help, as well as more psychological coping mechanisms to deal with stress (House et al. 1988; Lillard and Waite 1995; Mirowsky and Ross 2003). These resources are coupled with overall lower levels of stress from work and family problems; for example education is associated with lower levels of divorce (House, Landis, and Umberson 1988; Lantz et al. 2005).

A number of studies have estimated the relationship between education and mortality. This relationship is most often estimated by examining years of schooling categorized into a series of dummy variables (see Christenson and Johnson, 1995; Kitagawa and Hauser, 1973; Molla, Madans, and Wagener, 2004) or as a linear predictor (Elo, Martikainen, and Smith, 2006; Lynch, 2006; Preston and Elo, 1995; Zajacova, 2006). The use of a linear term for education is

extremely problematic, as much research has shown that there is not a monotonic relationship between years of education and mortality risk (Backlund, Sorlie, and Johnson 1999). Moreover, while the use of dummy variables skirts the issue of non-linearity in the relationship between education and mortality, it often obscures potential differences in the shape of the mortality curve as well as different threshold points in the data that may vary substantially between demographic populations. This is particularly true among Hispanic populations in the U.S. who have health outcomes that are substantially better than whites at similar socioeconomic statuses (SES) (Markides and Coreil 1986; Markides and Eschbach 2005).

Because there are often too few observations and deaths in each year of education, it is often impossible to use dummy variables for each year of education to estimate a mortality curve. The use of penalized spline on education allows us to use the all the information available and create stable confidence intervals for single years of education. We therefore can examine mortality curves across all years.

METHODS

Data and Measures

We employ data from the public-use National Health Interview Survey (NHIS) Linked Mortality Files (NHIS-LMF) from 1989 through 1996. NHIS-LMF links the 1989 through 1996 to mortality using the National Death Index (NCHS 2005, 2007). This data set is ideal to examine the relationship between education and morality because the data set is extremely large, nationally representative, and allows us to examine the link between education and mortality among several sub populations. To provide a public-use version of these data, NCHS perturbed the dates or cause of death for a select sample of records to ensure that individuals could not be

identified. Lochner et al. (2008) demonstrate that the public-use and restricted data sets produce equivalent results for both overall and cause-specific mortality, which perfectly suits our analysis. We restrict our sample to respondents over 25 years of age at the time of the survey and no older than 75 years of age[1] (N=670,571). We also restrict our analysis to respondents who identify as non-Hispanic white, non-Hispanic black, or Mexican (N=616,344) respondents.

The dependent variable for our analyses is dichotomous: survival versus death. Following Kom, Graubard, and Midthune (1997), we use age to indicate the time to death, which ensures that our mortality analyses are age-adjusted. Analyses are disaggregated by sex and race/ethnicity. We perform these analyses for non-Hispanic whites, non-Hispanic blacks, and Mexican respondents, by sex. Education is coded as a continuous variable that ranges from 0 years of education and it top-coded at 19 years of education. We control for nativity status in analysis of non-Hispanic black and non-Hispanic white respondents and disaggregate by nativity status among Mexican respondents. Marital status is included in the models as married (referent), widowed, divorced or separated, or never married. We also control for employment status as employed (referent), unemployed, and not in the labor force.

Method

We use Cox proportional hazard models with a penalized spline on education to model the non-linear relationship between education and all-cause mortality. We include the appropriate weights and strata to the model to ensure it's nationally representative character. Penalized splines estimate the non-linear relationship between education and morality by smoothing across a number of specified knots. This process begins with a large number of knots (splines) that fit

[1] We do this because at the oldest ages, educational outcomes and mortality estimates become unstable.

the data exceptionally well which are then penalized by λ which is a smoothing parameter. As λ becomes larger, the roughness penalty increases, smoothing the data. Conversely, as λ approaches zero, the roughness penalty becomes null, resulting in an over-fitting of the data. It is therefore important to choose a roughness penalty (theta) that both ensures an accurate representation of the data and does not reflect "high frequency type of variation that we would prefer to ignore". Theta can vary between 0 and 1, where 0 is as close to an exact fit of the relationship between education and mortality, and 1 is as smooth of a fit as possible. We first estimate the models based up on AIC information criteria where the degrees of freedom are chosen automatically. We then use our substantive knowledge to pursue more conservative models restricting the degrees of freedom to four and the theta to .9[2], this will ensure that our plots are smooth and reflect the real relationship between education and mortality and not just the intricacies of the data.

RESULTS

Descriptive Statistics

Table 1 provides descriptive information on the distribution of education across race/ethnic categories as well as the percent dead in each category.

(Table 1 about here)

Across race/ethnic groups we see that foreign born Mexican Americans have the highest percentages within the lowest education groups; 51% of males and 56% of females have 8 years or fewer of education, roughly twice the amount of U.S. born Mexican Americans. Both of these rates, however, are much higher than non-Hispanic white and non-Hispanic black respondents,

[2] Models with varying degrees of freedom and theta are available and will be provided for those who wish to examine them at the PAA meeting.

only around 6% of white males and females and 11% of black males and females have eight years or fewer of education. A similar gradient emerges at the higher ends of education. Among whites, 31% of males and 25% of females have between 15 and 19 years of education compared to 7% of foreign born Mexican males and females, 14% of U.S. born Mexican males and 11% of U.S. born Mexican females and 18% of black males and females. There does not appear to any consistent graded relationships between percent dead by education groups in the descriptive statistics. We therefore turn to the Cox proportional hazard models and the non-linear relationship between education and mortality.

Non-Linear Effects of Education on Mortality

We first present model fit statistics for the relationship between education and mortality for models that use education as a linear term and those with the penalized spline with a theta specified at .9. For every group, models with a the penalized spline we see improvements in the R square, except for US born Mexican males, where the R square remains the same. We now turn to the hazard models, which are presented in graphical form. Across the X axis are years of education. Across the Y axis is mortality risk. The point 0.0 on the Y axis represents the mean mortality risk for that population, thus, points above 0.0 represent increased mortality risk, and those below represent decreased mortality risk. For example, those with at the .2 mark are 1.22 standard deviations above the mean mortality risk, and those at .4 are 1.49 standard deviations above the mean mortality risk.

Females. The first four graphs present the mortality risk curves for females. There are indeed substantial differences between the four race/ethnic groups presented. Non-Hispanic white

females appear to have a somewhat linear relationship with mortality in that they experience the highest mortality risk at the lowest education level that moves in a downward trend to the highest education levels, between 0 and 12 year of education there is an above the mean mortality risk for the population. While there is an upward swing in mortality risk at the highest education group, it does not cross the 0.0 line, and therefore does not represent an actual increase in mortality risk compared to the mean risk.

(Panel A about here)

A very different trend emerges among non-Hispanic black females. A bifurcated relationship between education and mortality appears: females with 0 to roughly 8 years of education and females with greater than 15 years of education have a lower mortality risk than the average for the population, although there is a sharp decrease in mortality risk at the higher education levels. This suggests that far from being a linear relationship between years of education and mortality, those individuals who have fewer years of education may do as well as those with high school educations. A third unique non-linear appears in the relationship between education and mortality among Mexican women. While there appears to be no mortality penalty at the very lowest levels of education for foreign born Mexican women, the two curves look roughly similar and show a fairly flat relationship between education and mortality for about the first 14 years, after this point, there is a dramatic decrease in mortality risk at the highest ages.

Males. Among non-Hispanic white males, a trend similar to females emerges with the highest mortality risk occurring at the lowest education groups and an uneven progression downward to the lowest risk occurring for those with the highest years of education, with a sharp increase in

risk at the 19 years of education. This downward trend however is unique to white men and women.

(Panel B about here)

A bifurcated similar to non-Hispanic black females emerges among black males: those respondents with 0 to roughly 9 years of education and those with over 15 years of education have below the mean mortality risk for the population. It is among those respondents with 9 to 15 years of education that there is an increased mortality risk. A similar trend emerges among foreign born Mexican males where those persons with the lowest and highest years of education have decreased mortality risk than those with between 5 and 12 years of education. Amongst U.S. born Mexican males, however, there appears to be an almost flat relationship between education and mortality. There is a small increase in risk between 4 and 12 years, however, there is no major drop at either the high or the low ends of education, suggesting that while there may be slight benefit for attaining higher years of education and life expectancy, this gain is minimal, and as the confidence intervals show, not significant.

DISCUSSION AND CONCLUSION

The results presented in this table are the first to examine the non-linear relationship between education and mortality using a penalized spline, moreover, using the NHIS-LMF data we are able to examine the relationship by sex and a variety of race/ethnic groups. We find very different relationships between education and mortality across race/ethnic groups. Among non-Hispanic whites we find a general downward trend between education and mortality risk. This curve represents the general understanding of how education is related to mortality, that as one increases their education they decrease their mortality risk. This trend, however, appears to be

unique to white males and females. Indeed, very different curves emerge among non-Hispanic black and Mexican men and women. A series of bifurcated relationships emerges among non-whites, most dramatically among foreign-born Mexican males, and less dramatically among foreign-born Mexican females, and black males and females. Among U.S. born Mexican males and females, there appears to be almost no relationship between education and mortality except among women with greater than 15 years of education.

The bifurcated relationship between education and mortality among Mexicans can partially be explained by the Hispanic paradox, which has shown that Mexican Americans have health outcomes comparable to non-Hispanic whites in the United States despite the major SES differences between the two groups (Markides and Coreil 1986). While much research has sought to "solve" this paradox, the majority of research that has emerged in the last two decades remains supportive of the paradox, particularly among Mexican born males (Markides and Eschbach 2005). It is among this group that we see the best mortality outcomes among the least educated.

The mortality curve presented here across education for non-Hispanic black males and females, however, cannot be explained by the Hispanic paradox or migration. What appears may be partially explained by the "John Henryism" hypothesis. Drawing upon the tale of John Henry who beat engaged in a "steal driving" contest against a machine, and almost immediately after winning the contest, dropped dead from exhaustion. Applied to the health of non-Hispanic blacks in the U.S., this hypothesis suggests that "high effort coping" may lead to worse health outcomes, particularly cardiovascular disease risk (James 1994). Non-Hispanic blacks in the United States in general, have lower SES status than non-Hispanic whites that are coupled with high rates of discrimination that lead to overall higher levels of stress. Thus, those individuals

who do not engage in high effort coping, or attempt to move beyond their current SES status may not encounter the same levels of stress of those who do, helping to explain the decreased mortality risk at the low end of the education scale. At the high end, we also see a mortality benefit that graduate from college and see the benefits of their labor. It is among those individuals that only graduate from high school or attend some college but do not graduate where the greatest mortality risks emerge.

While the processes underlying the peculiar relationship between education and mortality among non-Hispanic blacks is in need of further exploration, what is clear is that the relationship is both unique and non-linear. In fact, the only group among which there appears to be any sort of linear relationship is non-Hispanic white men and women. This research highlights the importance of taking a closer look at the relationship between education and mortality, particularly by race/ethnicity. Moreover, it suggests that unless one graduates from college, higher levels of education may not be a universal solution to improving health across race/ethnic groups. More research is needed to understand why there are mortality benefits at the lowest years of education and varying levels of mortality benefits at the highest levels of education across groups.

Table 1. Descriptive Information on Mortality and Education by Sex and Race/Ethnicity

	Non Hispanic White		Non-Hispanic Black		Foreign Born Mexican		US Born Mexican	
	Males	Females	Males	Females	Males	Females	Males	Females
Yrs of Education								
0 to 2	0.36	0.23	1.70	0.97	10.21	10.34	4.17	4.72
3 to 5	0.87	0.63	3.24	2.16	15.70	17.70	5.56	5.76
6 to 8	5.23	4.53	7.88	7.10	25.29	28.04	12.35	13.14
9 to 11	9.49	9.84	16.12	17.46	14.69	13.58	15.40	16.72
12 to 14	52.51	59.86	52.95	55.12	22.82	23.32	4.83	48.51
15 to 17	21.45	18.54	12.64	12.59	4.27	4.49	10.17	8.44
18 to 19	10.09	6.37	5.47	4.61	3.01	2.53	4.06	2.71
% Dead by Education								
0 to 2	0.12	0.01	0.70	0.31	0.93	1.07	0.99	0.99
3 to 5	0.33	0.17	1.45	0.68	1.06	1.31	1.03	0.88
6 to 8	1.77	1.19	2.55	1.84	2.06	1.78	2.19	1.44
9 to 11	1.93	1.67	3.02	2.54	0.73	0.52	1.51	0.95
12 to 14	5.88	5.09	5.57	4.09	1.12	1.00	2.96	1.81
15 to 17	1.59	0.84	0.87	0.61	0.24	0.12	0.57	0.25
18 to 19	0.78	0.31	0.69	0.33	0.01	0.10	0.29	0.11
N	238,506	260,490	35,271	50,703	6,513	6,116	8,751	9,993

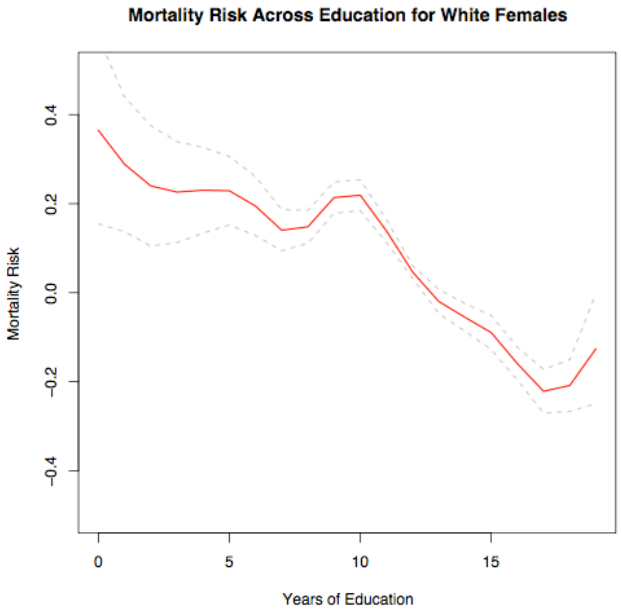
Source: Derived from NHIS-LMF.

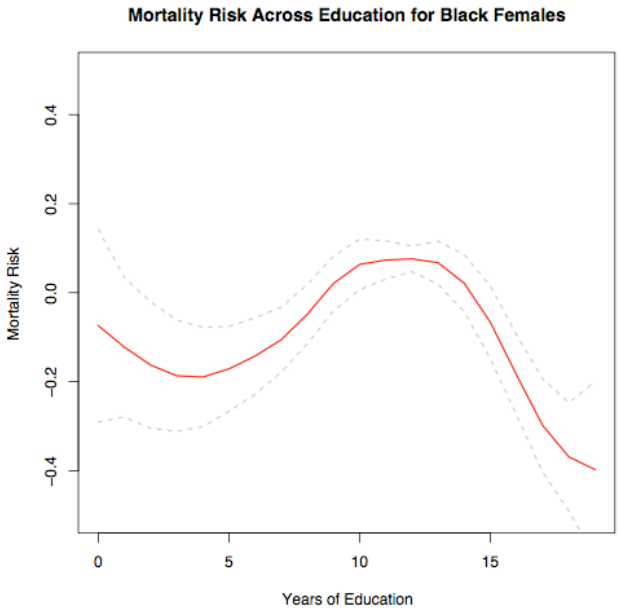
Table 2. Model Fit Statistics

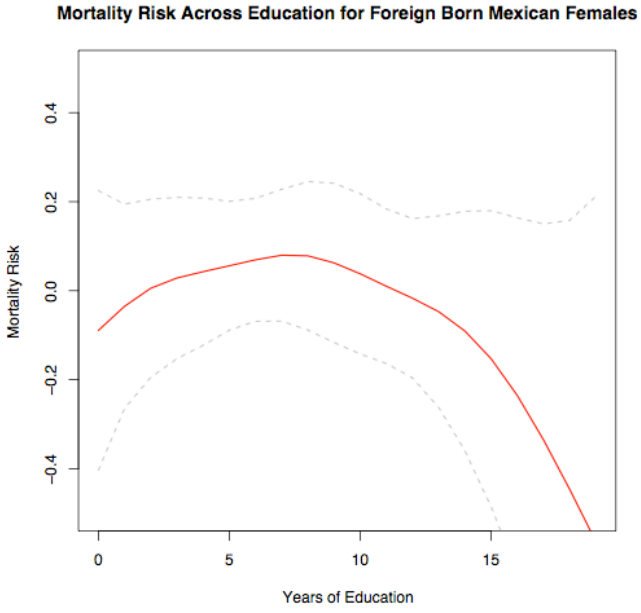
	Non Hispanic White		Non-Hispanic Black		Foreign Born Mexican		US Born Mexican	
	Males	Females	Males	Females	Males	Females	Males	Female
Education as Linear Term								
R Square	0.008	0.003	0.013	0.007	0.005	0.003	0.008	0.002
Likelihood Ratio Test	1994	796.7	465	374	30.31	18.79	65.93	22.72
Degree Freedom	9	9	9	9	8	8	8	8
P	0	0	0	0	0.0019	0.016	3.15E-11	0.004
Education with Spline								
R Square	0.009	0.003	0.015	0.009	0.007	0.004	0.008	0.003
Likelihood Ratio Test	2202	845	527	436	42.5	22.8	73.2	26.1
Degree Freedom	16.4	16	15	14.9	11.4	11.2	12.1	11.8
P	0	0	0	0	1.72E-05	0.0214	8.01E-12	0.009

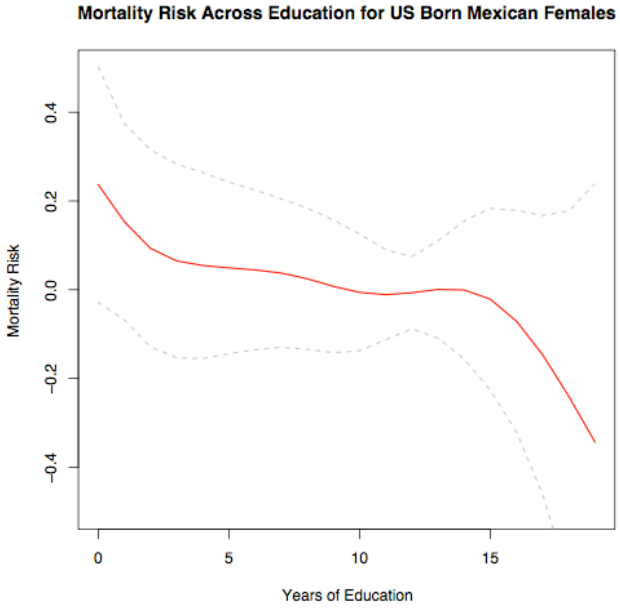
Source: NHIS-LMF

Panel A: Mortality Risk Curves for Females by Race/Ethnicity

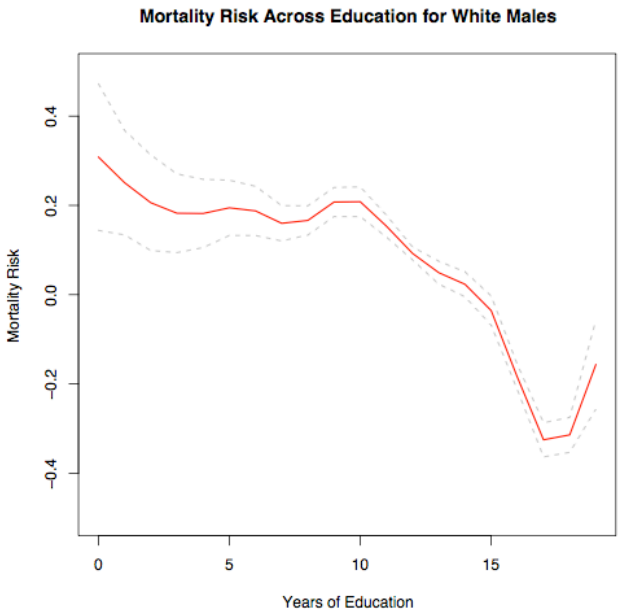


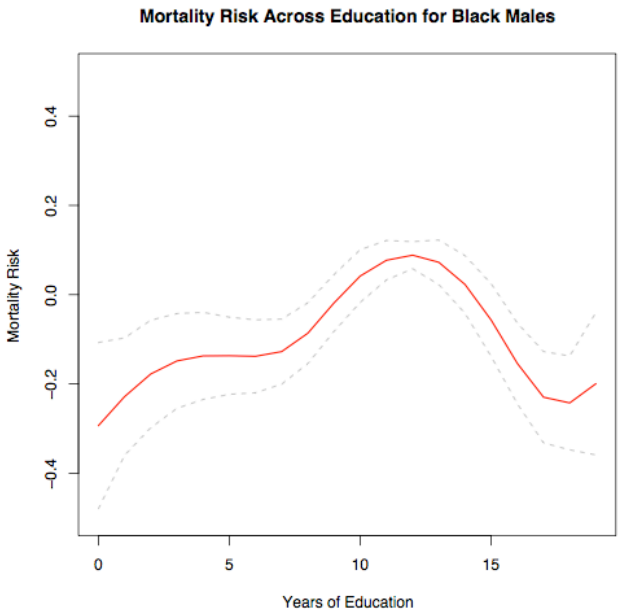


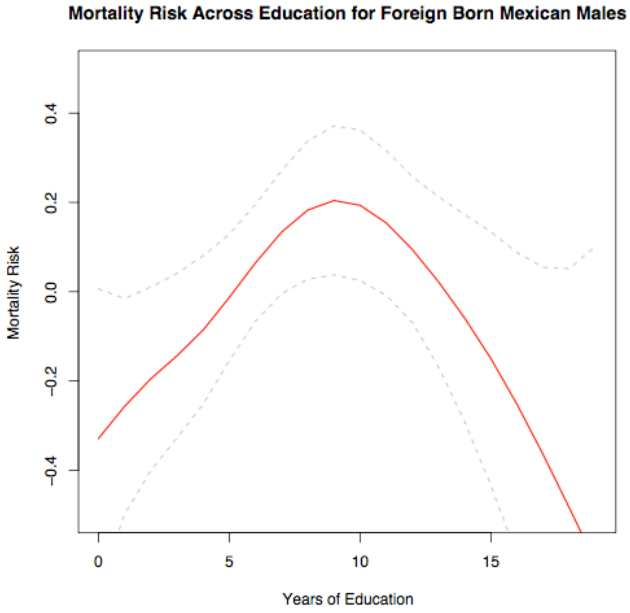


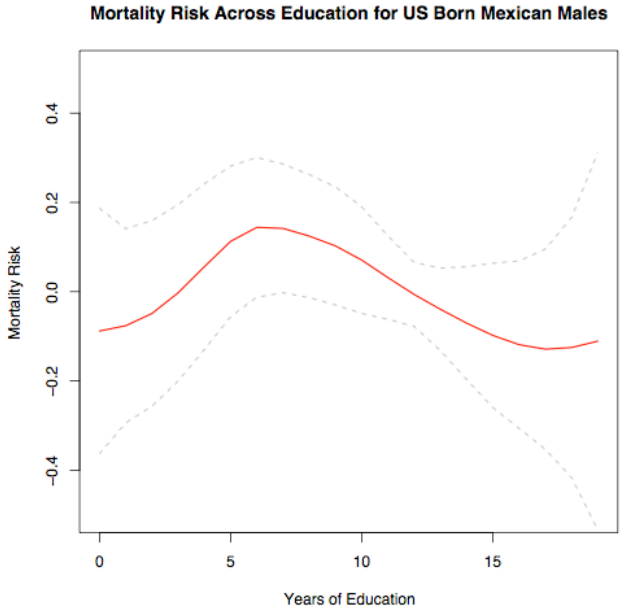


Panel B: Mortality Risk Curves for Males by Race/Ethnicity









REFERENCES

Backlund, Eric, Paul D. Sorlie, and Norman J. Johnson. 1999. "A Comparison of the Relationships of Education and Income with Mortality: The National Longitudinal Mortality Study." *Social Science and Medicine* 49:1373-84.

Becker, Gary S. 1964. *Human Capital: A Theoretical and Empirical Analysis, with Special Reference to Education.* New York, NY: Columbia University Press.

Christenson, Bruce A. and Nan E. Johnson. 1995. "Educational Inequality in Adult Mortality: An Assessment with Death Certificate Data from Michigan." *Demography* 32(2):215-30.

Elo, Irma T. and Samuel H. Preston. 1996. "Educational Differentials in Mortality: United States, 1979-85." *Social Science and Medicine* 42(1):47-57.

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

House, James S., Kristin R. Landis, and Debra Umberson. 1988. "Social Relationships and Health." *Science* 241:540-5.

House, James S., James M. Lepkowski, Ann M. Kinney, Richard P. Mero, Ronald C. Kessler, and A. Regula Herzog. 1994. "The Social Stratification of Aging and Health." *Journal of Health and Social Behavior* 35(3):213-234.

James, S. A. 1994. "John Henryism and the Health of African-Americans." *Culture, Medicine, and Psychiatry* 18: 163-82.

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

Kom, E., B.I. Graubard, and D. Midthune. 1997. "Time-to-Event Analysis of Longitudinal Follow-up of a Survey: Choice of the Time-Scale." *American Journal of Epidemiology*

145(1):72-80.

Lantz, Paula M., James S. House, James M. Lepkowski, David R. Williams, Richard P. Mero, and Jieming Chen. 1998. "Socioeconomic Factors, Health Behaviors, and Mortality: Results from a Nationally Prospective Study of U.S. Adults." *Journal of the American Medical Association* 279(21):1703-8.

Lantz, Paula M., James S. House, Richard P. Mero, and David R. Williams. 2005. "Stress, Life Events, and Socioeconomic Disparities in Health: Results from the Americans' Changing Lives Study." *Journal of Health and Social Behavior* 46:274-89.

Lillard, Lee and Linda Waite. 1995. "Til Death Do Us Part: Marital Disruption and Mortality." *American Journal of Sociology* 100:1131-56.

Link, Bruce and Jo Phelan. 1995. "Social Conditions and Fundamental Causes of Disease." *Journal of Health and Social Behavior* (extra issue):80-93.

Lynch, S. M. 2006. "Explaining Life Course and Cohort Variation in the Relationship between Education and Health: The Role of Income." *Journal of Health and Social Behavior* 47(4):324–38.

Markides, K. S. and K. Eschbach. 2005. "Aging, Migration, and Mortality: Current Status of Research on the Hispanic Paradox." *Journals of Gerontology: SERIES B* 60B(Special Issue II): 68-75.

Markides, K. S., and J. Coreil. 1986. "The Health of Southwestern Hispanics: An Epidemiologic Paradox." *Public Health Reports* 101:253-65.

Mirowsky, John and Catherine E. Ross. 1998. "Education, Personal Control, Lifestyle and Health: A Human Capital Hypothesis." *Research on Aging* 20(4):415-49.

-------. 2003. *Education, Social Status, and Health*. New York, NY: Aldine de Gruyter.

National Center for Health Statistics. Various years. *1997-2000 National Health Interview Surveys: Core Data*. Hyattsville, MD: U.S. Department of Health and Human Services.

-------. 2005. *The 1986-2000 National Health Interview Survey Linked Mortality Files: Matching Methodology*. Hyattsville, Maryland. (Available at: http://www.cdc.gov/nchs/data/datalinkage/matching_methodology_nhis_final.pdf).

-------. 2007. *National Health Interview Survey, Linked Mortality Public-Use Data File: 1986-2002 Survey Years*. Hyattsville, MD: U.S. Department of Health and Human Services.

Preston, Samuel H. and Irma T. Elo. 1995. "Are Educational Differentials in Adult Mortality Increasing in the United States?" *Journal of Aging and Health* 7(4):476-96.

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

Zajacova, Anna. 2006. "Education, Gender, and Mortality: Does Schooling Have the Same Effect on Mortality for Men and Women in the US?" *Social Science and Medicine* 63(8):2176-90.