CrossMark

Trends in Education-Specific Life Expectancy, Data Quality, and Shifting Education Distributions: A Note on Recent Research

Arun S. Hendi[1]

Abstract Several recent articles have reported conflicting conclusions about educational differences in life expectancy, and this is partly due to the use of unreliable data subject to a numerator-denominator bias previously reported as ranging from 20 % to 40 %. This article presents estimates of life expectancy and lifespan variation by education in the United States using more reliable data from the National Health Interview Survey. Contrary to prior conclusions in the literature, I find that life expectancy increased or stagnated since 1990 among all education-race-sex groups except for non-Hispanic white women with less than a high school education; there has been a robust increase in life expectancy among white high school graduates and a smaller increase among black female high school graduates; lifespan variation did not increase appreciably among high school graduates; and lifespan variation plays a very limited role in explaining educational gradients in mortality. I also discuss the key role that educational expansion may play in driving future changes in mortality gradients. Because of shifting education distributions, within an education-specific synthetic cohort, older age groups are less negatively selected than younger age groups. We could thus expect a greater concentration of mortality at younger ages among people with a high school education or less, which would be reflected in increasing lifespan variability for this group. Future studies of educational gradients in mortality should use more reliable data and should be mindful of the effects of shifting education distributions.

Keywords Mortality · Education · Gradients · Socioeconomic status · Health

✉ Arun S. Hendi
arun.hendi@duke.edu

[1] Population Research Institute and Department of Sociology, Duke University, 276 Soc/Psych Building, Box 90088, 417 Chapel Dr., Durham, NC 27708-0088, USA

🖄 Springer

Introduction

There has been a great deal of confusion in recent years on the subject of educational differences in U.S. mortality. Studies have variously concluded that since the 1990s, life expectancy declined among the least-educated white men and women (Olshansky et al. 2012); mortality increased among middle-aged whites, due partly to increases in poisonings among the less-educated (Case and Deaton 2015); white life expectancy increased among all education-sex groups except for the least-educated white women (Hendi 2015); and life expectancy declined among the least-educated white men and women, conditions vastly improved for less-educated blacks, and lifespan variation increased among high school graduates (Sasson 2016). Demographers whose special-izations lie outside the area of mortality have been left wondering what's actually going on, while nonacademics have understandably been convinced by news reports that mortality conditions are dramatically worsening. The discrepancies among these dif-ferent conclusions is partly the result of faulty data or problematic methods in some of these studies. This comment tries to make sense of these radically different conclusions by replicating the analyses of the most recent of these articles—specifically, Sasson (2016)—using a different, more reliable data set.

Sasson (2016) reported that life expectancy declined for white men and women with less than a high school education and increased only modestly for white high school graduates, that conditions improved rapidly for less-educated blacks, and that lifespan variation increased for high school completers. In this comment, I argue that most of these novel findings are subject to a well-known numerator-denominator bias in the data used in Sasson (2016) and that these findings do not hold up when using a more reliable data set. In contrast to the Sasson estimates, I show that between 1990 and 2009, life expectancy increased or stayed constant for every race-sex-education group except for non-Hispanic white women with less than a high school education, a group whose relative and absolute size more than halved over this period. There was a robust increase in life expectancy among white high school graduates but a smaller improve-ment among black women who completed high school. There was no appreciable increase in lifespan variation for high school graduates, and most groups saw a decrease. I also point out that because of the rapid change in the education distribution over this period, the Sasson estimates may not necessarily be indicative of worsening socioeconomic gradients in mortality.

Dual Data-Source Bias

Sasson (2016) constructed education-specific mortality rates using National Vital Statistics System (NVSS) death certificate data on the number of deaths by education (numerator) and census data on the midyear population by education (denominator). Because self-reports of education in the census differ systematically from funeral director/family-reported education on death certificates, these education-specific death rates are subject to dual data-source or numerator-denominator bias. In 1989, 38 % of decedents who were listed as high school graduates on their death certificates previ-ously self-reported being high school noncompleters (Sorlie and Johnson 1996). That number is close to 23 % for decedents in 1992–1998 (Rostron et al. 2010). Thus, not

only is differential educational classification a problem, but the bias changes over time. Furthermore, the problem is not restricted to the less-educated. Of decedents who previously self-reported having more than a high school education, only 71 % of whites and 63 % of blacks were reported as having an education beyond high school listed on their death certificates (Rostron et al. 2010). The misclassification problem is thus equally severe at the upper end of the education distribution and varies by race. Dual data-source bias leads to inaccurate education-specific death rates, and the degree of error varies by time, age, race, and sex.

This bias is mentioned in Sasson (2016), but the article does not assess the sensitivity of the estimates to the bias and does not use available survey-linked mortality data that would allow one to avoid this problem.

Replication Using NHIS

I replicate the Sasson (2016) analyses using data from the Integrated Health Interview Series (IHIS) version of the National Health Interview Survey (NHIS) 1986–2009 linked to mortality follow-up through 2011 (MPC 2015). The NHIS is an annual survey of the U.S. civilian noninstitutionalized population and is the largest mortality-linked American survey that allows for the estimation of mortality for the periods in this study. Although it does not include people housed in nursing homes or the incarcerated and may miss people living on the margins of society, it provides a distinct advantage over vital statistics: it allows one to compute education-specific death rates that are unaffected by dual data-source bias because education is self-reported by the individual at the time of survey, so differential educational classification is not a problem. In other words, the NHIS allows one to estimate consistent education-specific life expectancies for the civilian noninstitutionalized population, whereas the NVSS/census data do not allow estimation of consistent education-specific life expectancies for *any* population.

Respondents aged 25 and older at baseline are allowed to contribute at most 10 person-years of exposure in order to reduce bias resulting from institutionalization of the baseline population (see Online Resource 1). I compute age-sex-education–specific death rates for the non-Hispanic white and black populations using standard occurrence-exposure ratios for five-year age groups (25–29, . . . , 85+). I construct period life tables for three periods: 1988–1992 (centered on 1990), 1998–2002 (2000), and 2007–2011 (2009), computing graduated $_na_x$ values (Keyfitz 1966). Sample sizes are provided in Table 3 of the appendix. Sasson (2016) reported three main analyses, each of which I replicate in this comment: life expectancy at age 25 (e_{25}), standard deviation of person-years lived beyond age 25 (S_{25}), and an approximate decomposition of the Kullback-Leibler divergence (KLD) into mean and variance components.

Results

Tables 1 and 2 present NHIS-based estimates of e_{25} and S_{25}, respectively, and can be compared with Sasson's second and third tables (2016: tables 2 and 3). Because the NHIS is representative of the noninstitutionalized population, the NHIS-based e_{25} estimates are higher than vital statistics–based estimates. I focus on changes between

Table 1 Life expectancy at age 25 by race, sex, and education; United States, 1990–2009

	Non-Hispanic Whites						Non-Hispanic Blacks[a]					
	Females			Males			Females			Males		
Education	1990	2000	2009	1990	2000	2009	1990	2000	2009	1990	2000	2009
Less Than High School	54.3	53.1	51.9	46.2	47.8	47.2	50.0	49.4	50.7	41.6	44.5	45.7
High School	57.8	57.6	58.8	50.5	51.9	53.1	53.6	53.9	54.0	45.9	48.4	50.4
Some College	60.1	59.3	61.1	51.7	52.8	54.4	57.8	56.1	57.6	46.1	49.2	54.2
College or More	60.8	60.8	63.4	54.4	56.6	59.2	56.3	57.9	60.7	53.9	52.6	58.7
Total	57.6	57.5	59.4	50.3	52.4	54.4	52.3	53.3	55.1	44.5	47.9	50.7

Source: Author's calculations based on NHIS 1986–2009 with mortality follow-up through 2011.

[a] Estimates for non-Hispanic blacks with some college or more are less reliable.

1990 and 2009 because this comparison allows the detection of significant changes and because of the interest here in secular trends.

Several differences between the NHIS-based estimates and Sasson's estimates are immediately apparent. Sasson's finding that e_{25} declined by 3.1 years and 0.6 years for white women and men (respectively) with less than a high school education is not supported by the NHIS estimates. The NHIS-based estimates indicate that from 1990 to 2009, e_{25} declined by 2.5 years for white women and *increased* by 1.1 years for white men with less than a high school education. In addition, there are discrepancies between the Sasson (2016) and NHIS estimates of change in e_{25} between 1990 and 2010/2009 for the following groups (with the Sasson minus the NHIS estimates in parentheses): college or more (+1.1 years) for white women; less than high school (−1.6 years) and high school (−0.8 years) for white men; less than high school (+1.2

Table 2 Standard deviation of life expectancy at age 25 (S_{25}) by race, sex, and education: United States, 1990–2009

	Non-Hispanic Whites						Non-Hispanic Blacks[a]					
	Females			Males			Females			Males		
Education	1990	2000	2009	1990	2000	2009	1990	2000	2009	1990	2000	2009
Less Than High School	16.6	14.8	16.1	15.8	14.3	15.9	18.2	17.0	16.6	16.7	16.5	15.8
High School	14.7	13.7	14.2	14.3	13.7	14.9	16.4	16.0	15.3	15.5	15.5	15.5
Some College	14.9	13.2	13.8	14.8	13.8	14.2	21.1	15.4	15.4	15.7	16.1	15.6
College or More	14.3	12.3	12.2	13.1	12.8	11.7	13.2	12.4	14.6	20.7	14.7	13.5
Total	14.9	13.2	13.5	14.2	13.4	13.9	16.9	15.2	15.1	15.8	15.2	15.4

Source: Author's calculations based on NHIS 1986–2009 with mortality follow-up through 2011.

[a] Estimates for non-Hispanic blacks with some college or more are less reliable.

years), high school (+3.1 years), and some college (+4.8 years) for black women; and less than high school (+1.8 years) and college or more (+2.8 years) for black men. The Sasson estimates of the widening of the gradient exceed the NHIS estimates by 1.7 years for white women, 2.1 years for white men, and 1.0 years for black men.

The NHIS estimates do not indicate a dramatic increase in lifespan variability for any group.[1] Among high school graduates, for whom Sasson found the greatest evidence of increasing lifespan variability, S_{25} increased by 0.6 years for white men, declined for both black and white women, and remained unchanged for black men.

Figures 1–4 plot decompositions of the KLD, where the reference groups are white college graduates in 2009 (comparable with Sasson 2016: figs. 2–5). The NHIS estimates indicate that variance explains relatively little of the difference in age at death distributions between white college graduates in 2009 and other groups. The variance component exceeds 50 % of the total KLD in only 8 of 46 cases. Five of these eight cases are black women or men with some college or more, groups for which estimates are less reliable. The remaining three cases are white women with some college or more, a group with exceptionally low mortality and a very similar mortality profile to the reference group. Typically, variance explains only 20 % of the total KLD. No significant increases in the variance component occurred over time for any race-sex-education group.

Discussion

The central point of this article is that studies using dual data sources to compute education-specific mortality tend to overstate mortality among the less-educated, mistakenly report long-term declines in life expectancy among less-educated whites, overstate educational differences in mortality, and exaggerate the convergence in mortality between less-educated blacks and whites. The discrepancies between the trend estimates in Sasson (2016) and this comment are sizable and likely due to numerator-denominator bias in the NVSS/census data. Although I cannot definitively rule out that the two sets of results are discrepant as a result of the NHIS exclusion of the institutionalized, a simulation shown in Online Resource 1 indicates that this is unlikely. What is known definitively on the basis of past studies is that the Sasson estimates suffer from dual data-source bias. The NHIS allows the estimation of a time series of reasonably accurate education-specific death rates for the noninstitutionalized population. The NVSS/census data do not allow the estimation of accurate education-specific death rates for *any* population.

Although Sasson (2016) described worsening life expectancy for white high school noncompleters, only modest increases in life expectancy among white high school graduates, and rising lifespan variability among black and white high school graduates, those findings were based on flawed data. Estimates based on more reliable data lead to different conclusions. Between 1990 and 2009, life expectancy improved or stayed constant for every race-sex-education group except for non-Hispanic white women

---

[1] These results should be interpreted with caution given that overall lifespan variability increased by a slight amount for non-Hispanic whites in U.S. life tables but decreased slightly in NHIS data.
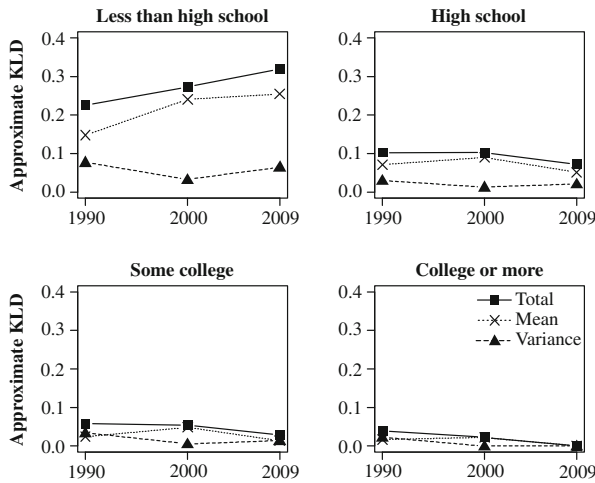
Fig. 1 Divergence and convergence in age-at-death distribution by education category for non-Hispanic white women. The reference category is non-Hispanic white women with a college degree or more in 2009. *Source*: Author's calculations based on NHIS 1986–2009 with mortality follow-up through 2011

with less than a high school education, consistent with previous research (Hendi 2015). White high school graduates experienced robust increases in life expectancy on the order of 1.0–2.6 years. Black women with a high school education did not experience life expectancy increases that were as large as indicated in Sasson, which raises doubts about that article's conclusion that there is a black-white convergence in life expectancy for the less-educated. Lifespan variability does not appear to have risen dramatically among high school graduates between 1990 and 2009. The magnitudes of changes in S_{25} aren't large, and the relative magnitudes of the variance components of the KLD are
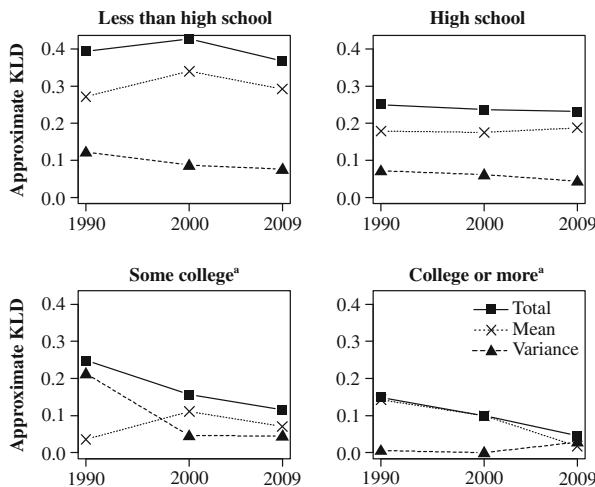

Fig. 2 Divergence and convergence in age-at-death distribution by education category for non-Hispanic black women. The reference category is non-Hispanic white women with a college degree or more in 2009. [a]Estimates are less reliable for black women with some college or more. *Source*: Author's calculations based on NHIS 1986–2009 with mortality follow-up through 2011
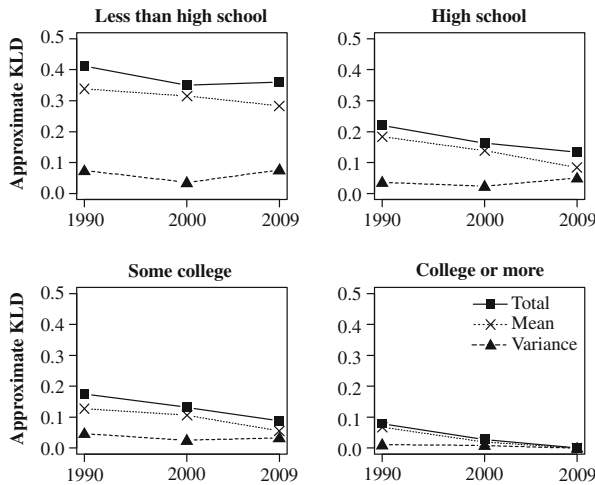
Fig. 3 Divergence and convergence in age-at-death distribution by education category for non-Hispanic white men. The reference category is non-Hispanic white men with a college degree or more in 2009. *Source*: Author's calculations based on NHIS 1986–2009 with mortality follow-up through 2011

quite small. Indeed, lifespan variability explains only a small fraction of the difference in mortality variation between less-advantaged groups and white college graduates in 2009. Trends in the KLD over time appear to be driven mostly by changes in the mean rather than the variance.

These findings suggest that researchers should be more cautious in making far-reaching conclusions about future mortality scenarios and trends in educational gradients in mortality. It is quite a leap to conclude that a slightly higher S_{25} means worse future mortality conditions over the long term. Although the vital statistics/
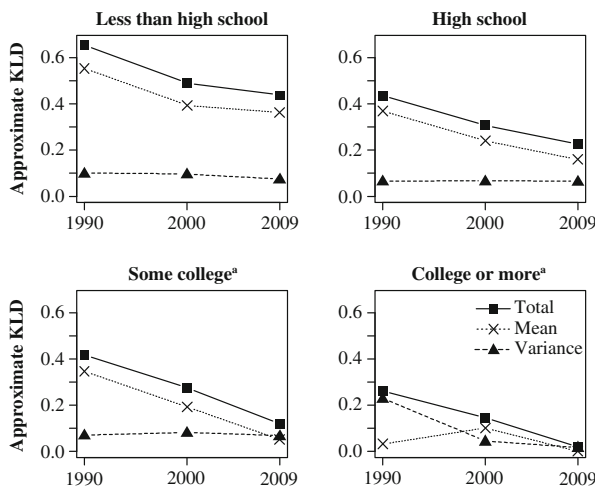
Fig. 4 Divergence and convergence in age-at-death distribution by education category for non-Hispanic black men. The reference category is non-Hispanic white men with a college degree or more in 2009. [a]Estimates are less reliable for black men with some college or more. *Source*: Author's calculations based on NHIS 1986–2009 with mortality follow-up through 2011

census rates used in Sasson (2016) have lower sampling variance, their 20 % to 40 % bias is unacceptable, especially when the same results are not obtained from a different, more reliable data set. In short, the evidentiary basis for such conclusions is weak.

What does this mean more broadly for the literature on education and mortality? For consumers of this literature, it means that we shouldn't trust NVSS/census-based estimates of education-specific mortality. Sometimes this has led to erroneous conclusions, as in Sasson (2016) and Olshansky et al. (2012). In other cases, some conclusions hold true: in Case and Deaton (2015), the conclusion that poisoning mortality is concentrated among the less-educated remains correct, but numerator-denominator bias leads to inconsistent estimates of the magnitude of education-cause-specific mortality (see Ho 2017).[2] Here is what we do know: to the best of our knowledge, life expectancy appears to have improved or stagnated since 1990 for all race-sex groups except for non-Hispanic white women. Race differences in life expectancy remain sizable, although they have narrowed since 1990. The difference in life expectancy between the highest and lowest education categories is growing, in part due to the increasing negative selectivity of the lowest education category (discussed later herein).

Mortality researchers need to adhere to a stronger set of standards and not countenance the use of data known to be faulty. And unusual results should be questioned. Researchers should follow a set of guidelines that includes at least the following: age-standardize when the age interval is larger than five years; benchmark estimates against U.S. life tables to ensure they are ballpark-correct; report how $_na_x$ values are computed; and validate model assumptions, such as proportional hazards.

Although these and future studies of education-specific mortality should eschew dual data sources in favor of mortality-linked survey data, such as NHIS, there is still the question of what to do when researchers need larger sample sizes or want to study causes of death not clearly identified in surveys. One solution is to use formal demographic relations combining information from multiple data sets, thus avoiding numerator-denominator problems. Better data are needed going forward to study racial and socioeconomic variation in mortality. Such a data set would require, at a minimum, much larger sample sizes than the NHIS, the inclusion of institutionalized persons, and self-reported age, race, and education. Ideal candidates are the census and American Community Survey (ACS) linked to the National Death Index. In the meantime, the conclusions of studies that suffer from dual data-source bias should be discounted, particularly when their conclusions are at odds with established findings and trends.

Even though S_{25} for high school graduates hasn't increased dramatically in the NHIS, an increase in the future should not be surprising. The reason is compositional change, which is the tendency for the same education category to represent a lower segment of the SES spectrum for younger cohorts than for older cohorts (Hendi 2015). Compositional change has been a focus of recent research on trends in educational gradients in life expectancy (Bound et al. 2015; Dowd and Hamoudi 2014; Hendi 2015), and it is clear it can also influence lifespan variation. The category "high school

---

[2] However, the reported increase in mortality among middle-aged white men in Case and Deaton (2015) may be overstated because of failure to age-standardize.

graduate" makes up a smaller component of each successive birth cohort, leaving a smaller and more select group at the younger ages and a larger and less-select group at the older ages. In a cross-section of time (i.e., a synthetic cohort), the proportion of younger people with high school or less is much smaller than the equivalent proportion for older people. For example, among white women in 2010, 31 % of 25- to 29-year-olds have a high school education or less, compared with 74 % of 85- to 89-year-olds. These younger women cannot be said to have the same status as the older women in the same education group. The younger women are more negatively selected than the older women, and as these cohorts age, we will observe an increasingly negatively select set of women at the younger ages relative to the older ages. This would likely lead to a greater concentration of mortality at the younger adult ages, which is precisely what is being captured by increasing lifespan variability.

Compositional change further affects our interpretation of socioeconomic gradients in period life expectancy by pairing young, low-status individuals with older, relatively higher-status individuals in the same synthetic cohort. Thus, a period life expectancy for people with a high school education cannot be said to correspond to any one level of status. Furthermore, changes over time in educational differences in life expectancy do not necessarily correspond to changes in the socioeconomic status (SES) gradient in life expectancy, given that the two extreme education categories represent different mixtures of SES levels at different points in time. Compositional change is important for education-specific life expectancy because it can lead to synthetic cohort estimates that distort age patterns of mortality and thus inflate lifespan variation and invalidate our primary measure of SES gradients in mortality.

How should researchers interpret a potential increase in lifespan variability or a slowdown of life expectancy gains for a low-education group in the context of compositional change? Does it signal a dramatic shift in the mortality regime and a bleak future, as suggested by Sasson (2016)? I think the answer is that researchers should be cautious. Different data sets give different answers, and the best available data on education-specific mortality suggest that the future may not be all that bleak.

None of this is to say that there is no reason for concern about American mortality. The United States is currently among the lowest-ranking developed countries in terms of life expectancy, and much of that is due to our uncharacteristically high mortality at the younger adult ages, with deaths from drug overdose, motor vehicle accidents, and homicide being more frequent than in other countries (Ho 2013). New estimates suggest that American life expectancy has stagnated in recent years and may even have declined between 2014 and 2015 (Xu et al. 2016). Gains in U.S. life expectancy are distributed unevenly across social groups, with the more-advantaged sweeping up the majority of the gains (Elo 2009; Hendi 2015). Questions about socioeconomic variation in mortality are important. However, using data subject to a 20 % to 40 % measurement error can lead to faulty conclusions. Mortality researchers should take greater care in the future to use high-quality data and account for the effects of changing educational composition.

Appendix

Table 3 Sample sizes for life table estimation by race, sex, and education for ages 25+: National Health Interview Survey, 1986–2009 linked to mortality follow-up through 2011

	White Males		White Females		Black Males		Black Females	
	Deaths	PY	Deaths	PY	Deaths	PY	Deaths	PY
1990								
LTHS	3,331	102,812	2,777	120,469	909	28,752	805	40,678
HS	2,143	192,178	2,129	258,912	289	27,772	297	43,271
SC	823	99,175	721	117,248	109	12,504	103	18,917
C+	874	137,830	501	112,542	61	8,081	66	11,720
2000								
LTHS	4,954	133,521	5,385	154,252	1,271	36,248	1,529	53,837
HS	4,894	337,545	5,583	433,047	655	55,568	784	79,519
SC	2,337	214,717	2,164	261,459	282	31,227	301	48,522
C+	2,327	280,755	1,397	246,797	160	19,658	175	28,577
2009								
LTHS	2,179	61,199	2,391	66,600	668	19,469	795	28,944
HS	3,309	215,157	3,792	248,632	573	46,743	687	59,245
SC	2,131	183,148	2,029	223,476	297	34,558	352	53,561
C+	1,834	217,312	1,112	210,783	131	20,385	165	29,339

Notes: All estimates refer to the analytic (unweighted) sample. PY stands for number of person-years of exposure in the given period; LTHS, for "less than high school"; HS, for "high school"; SC, for "some college"; and C+, for "college or more." The tabulations refer to the non-Hispanic black and white populations.

References

Bound, J., Geronimus, A. T., Rodriguez, J. M., & Waidmann, T. A. (2015). Measuring recent apparent declines in longevity: The role of increasing educational attainment. *Health Affairs, 34,* 2167–2173.

Case, A., & Deaton, A. (2015). Rising morbidity and mortality in midlife among white non-Hispanic Americans in the 21st century. *Proceedings of the National Academy of Sciences, 112,* 15078–15083.

Dowd, J. B., & Hamoudi, A. (2014). Is life expectancy really falling for groups of low socio-economic status? Lagged selection bias and artefactual trends in mortality. *International Journal of Epidemiology, 43,* 983–988.

Elo, I. T. (2009). Social class differentials in health and mortality: Patterns and explanations in comparative perspective. *Annual Review of Sociology, 35,* 553–572.

Hendi, A. S. (2015). Trends in U.S. life expectancy gradients: The role of changing educational composition. *International Journal of Epidemiology, 44,* 946–955.

Ho, J. Y. (2013). Mortality under age 50 accounts for much of the fact that US life expectancy lags that of other high-income countries. *Health Affairs, 32,* 459–467.

Ho, J. Y. (2017). The contribution of drug overdose to educational gradients in life expectancy in the United States, 1992–2011. *Demography.* Advance online publication. doi:10.1007/s13524-017-0565-3

Keyfitz, N. (1966). A life table that agrees with the data. *Journal of the American Statistical Association, 61,* 305–312.

MPC (2015). *Integrated Health Interview Series: Version 6.12* [Minnesota Population Center and State Health Access Data]. Minneapolis: University of Minnesota. Retrieved from http://www.ihis.us

Olshansky, S. J., Antonucci, T., Berkman, L., Binstock, R. H., Boersch-Supan, A., Cacioppo, J. T., . . . Rowe, J. (2012). Differences in life expectancy due to race and educational differences are widening, and many may not catch up. *Health Affairs, 31*, 1803–1813.

Rostron, B. L., Boies, J. L., & Arias, E. (2010). *Education reporting and classification on death certificates in the United States* (Vital and Health Statistics, Series 2 No.151). Washington, DC: National Center for Health Statistics.

Sasson, I. (2016). Trends in life expectancy and lifespan variation by educational attainment: United States 1990–2010. *Demography, 53,* 269–293.

Sorlie, P. D., & Johnson, N. J. (1996). Validity of education information on the death certificate. *Epidemiology, 7,* 437–439.

Xu, J., Murphy, S. L., Kochanek, K. D., & Arias, E. (2016). *Mortality in the United States, 2015* (NCHS Data Brief No. 267)*.* Hyattsville, MD: National Center for Health Statistics.