Online Appendix: Mortality Change Among Less Educated Americans (Paul Novosad, Charlie Rafkin, Sam Asher)

A Appendix: Additional Tables and Figures

Table A1

Distribution of Deaths by Cause, Ages 25–69 (2018)

Cause of Death	Share of Total Deaths
Cancer	28.28
Heart and other diseases of the circulatory system	21.75
Poisoning, suicide, chronic liver disease ("deaths of despair")	11.82
Diseases of the respiratory system	8.00
Accidents and injuries (primarily falls, motor vehicles, assaults)	5.79
Endocrine, nutritional and metabolic diseases	5.36
Diseases of the nervous system	3.69
Cerebrovascular diseases	3.33
Infectious and parasitic diseases	2.72
Diseases of the digestive system	2.48
Diseases of the genitourinary system	1.95
Mental and behavioural disorders	1.92
Deaths not elsewhere classified	0.95
Diseases of the blood and immune disorders	0.88
Diseases of the musculoskeletal system and connective tissue	0.53
Congenital malformations, deformations and chromosomal abnormalities	0.27
Diseases of the skin and subcutaneous tissue	0.19
Pregnancy, childbirth and the puerperium	0.07
Diseases of the eye, ear, mastoid and adnexa	0.01
Certain conditions originating in the perinatal period	0.00

Note: The table shows the distribution of causes of death for individuals aged 25–69 in 2018. Categories are defined by major headings in the ICD-10 Cause-Of-Death lists. The categories of cancer, heart disease and deaths of despair are pooled across subcategories following the previous literature. See Appendix B for additional details.

Table A2

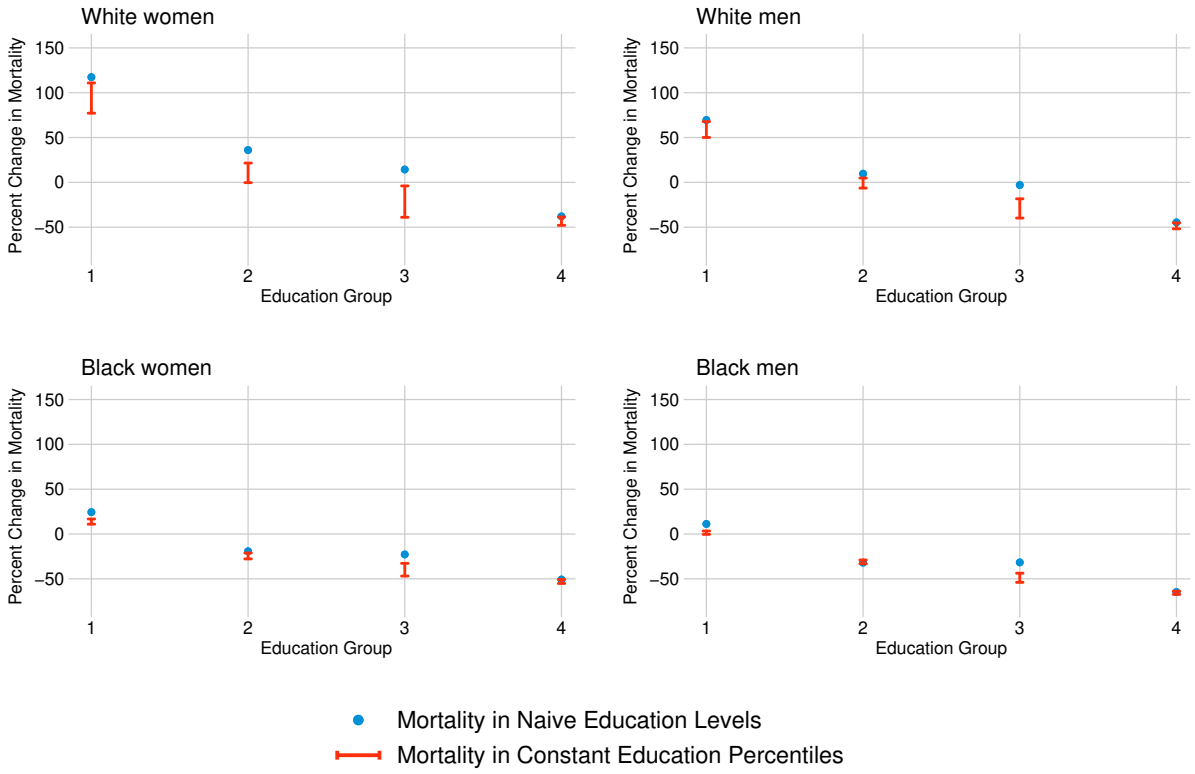
Age-Adjusted Changes in Mortality by Education Percentile and Cause
Ages 25–69, 1992–1994 to 2016–2018

		Injuries	Cancer	Heart Disease	Despair	Other	Total
White non-Hispanic Women							
Education Percentile	0–10	(+62%, +79%)	(+25%, +35%)	(+3%, +38%)	(+526%, +585%)	(+111%, +167%)	(+77%, +111%)
	10–45	(+15%, +36%)	(-26%, -18%)	(-38%, -16%)	(+274%, +354%)	(+18%, +52%)	(-0%, +21%)
	45–70	(-29%, +21%)	(-48%, -37%)	(-59%, -28%)	(+64%, +225%)	(-26%, +26%)	(-39%, -4%)
	70–100	(-50%, -34%)	(-49%, -47%)	(-61%, -55%)	(-14%, +42%)	(-38%, -28%)	(-48%, -39%)
White non-Hispanic Men							
Education Percentile	0–10	(+32%, +47%)	(+17%, +27%)	(+1%, +13%)	(+241%, +267%)	(+70%, +102%)	(+50%, +68%)
	10–45	(+2%, +16%)	(-30%, -24%)	(-38%, -31%)	(+156%, +186%)	(+3%, +16%)	(-6%, +5%)
	45–70	(-22%, +15%)	(-52%, -42%)	(-58%, -45%)	(+78%, +148%)	(-36%, -14%)	(-40%, -18%)
	70–100	(-44%, -35%)	(-55%, -52%)	(-63%, -59%)	(+17%, +39%)	(-54%, -48%)	(-52%, -45%)
Black non-Hispanic Women							
Education Percentile	0–10	(+11%, +19%)	(-9%, -6%)	(-16%, -11%)	(+152%, +168%)	(+23%, +32%)	(+11%, +17%)
	10–45	(-33%, -24%)	(-31%, -28%)	(-42%, -37%)	(+44%, +65%)	(-21%, -12%)	(-28%, -21%)
	45–70	(-49%, -25%)	(-45%, -40%)	(-60%, -49%)	(-8%, +48%)	(-38%, -20%)	(-47%, -33%)
	70–100	(-60%, -51%)	(-49%, -48%)	(-67%, -64%)	(-41%, -22%)	(-51%, -45%)	(-55%, -51%)
Black non-Hispanic Men							
Education Percentile	0–10	(+28%, +39%)	(-22%, -19%)	(-12%, -10%)	(+109%, +118%)	(-8%, -4%)	(-0%, +3%)
	10–45	(-18%, -8%)	(-46%, -44%)	(-37%, -33%)	(+30%, +40%)	(-36%, -32%)	(-33%, -29%)
	45–70	(-48%, -21%)	(-63%, -57%)	(-56%, -49%)	(-8%, +22%)	(-54%, -45%)	(-54%, -44%)
	70–100	(-69%, -63%)	(-68%, -66%)	(-66%, -63%)	(-45%, -34%)	(-69%, -66%)	(-67%, -64%)

Note: The table shows age-adjusted percentage change in mortality from 1992–1994 to 2016–2018 for individuals aged 25 to 69, by race, gender and education percentile bin. Each table entry shows the upper and lower bound on the percentage change in mortality over the sample period for the given cause and population subgroup. Ages are adjusted with a standardized U.S. population distribution, which holds constant that age distribution of the population across all years. Education percentile bins approximately describe the 2003 distribution of education across four categories: high schools dropouts (percentiles 0–10), high school graduates (10–45), some college (45–70) and B.A. or higher (70–100).

empty

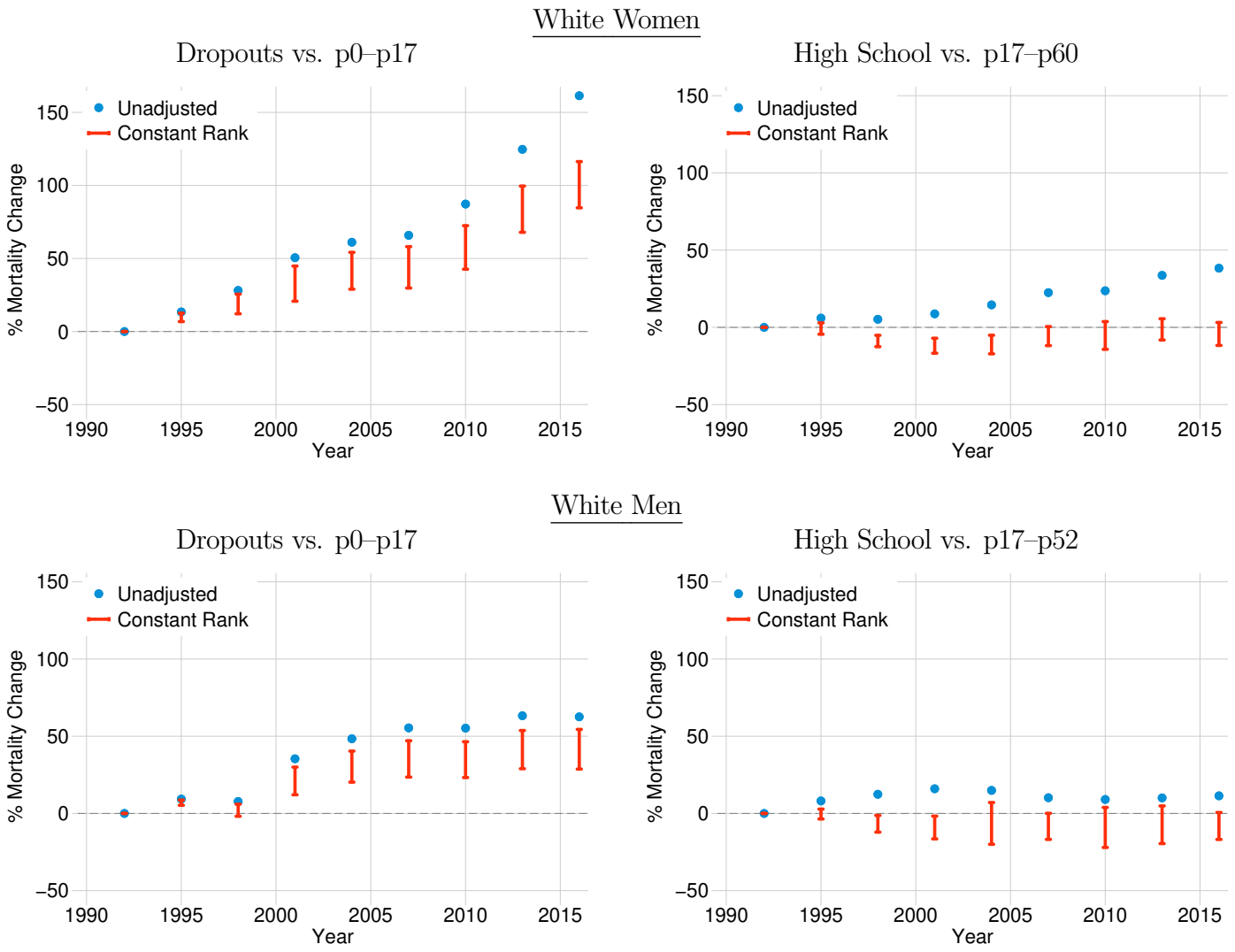
Figure A1
Bounds on Constant Percentile Mortality Change and Naive Point Estimates
Age-Adjusted Populations, 1992–1994 to 2016–2018

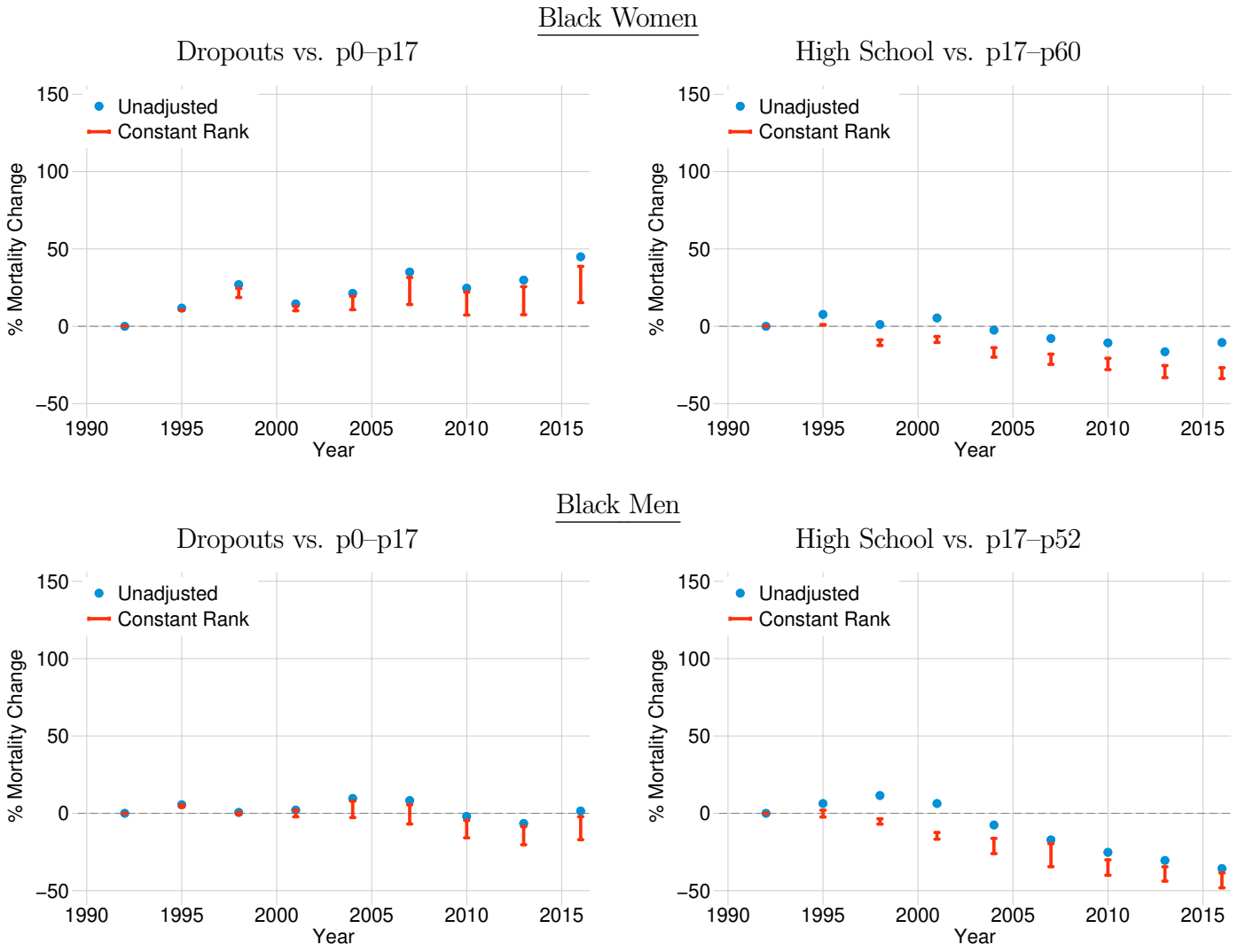


Note: "White" refers to non-Hispanic white and "Black" to non-Hispanic black. The line segments in the graph show bounds on age-adjusted mortality change from 1992–1994 to 2016–2018 for a standardized population, as in Table 2. The four education groups represent education percentiles (i) 0–10; (ii) 10–45; (iii) 45–70; and (iv) 70–100. The points in the graph show naive estimates of mortality rates at fixed education *levels*; the four education levels for the points represent (i) high school dropouts; (ii) high school completers; (iii) some college; and (iv) a B.A. or higher.

empty

Figure A2
Bounds on Constant Percentile Mortality Change and Naive Point Estimates:
50–54-year-olds, 1992–1994 to 2016–2018





Note: "White" refers to non-Hispanic white and "Black" to non-Hispanic black. The line segments in the graph show bounds on age-adjusted mortality change from 1992–1994 to 2016–2018 for 50-year-olds, separately by race and gender. The four education groups represent education percentiles (i) 0–10; (ii) 10–45; (iii) 45–70; and (iv) 70–100. The points in the graph show naive estimates of mortality rates at fixed education *levels*; the four education levels for the points represent (i) high school dropouts; (ii) high school completers; (iii) some college; and (iv) a B.A. or higher.

B Appendix: Data Construction

This section provides additional details on data construction. The All Cause Mortality file provided by the National Center for Health Statistics reports the number of deaths, by age, race, gender, education and state.

Mortality data. Beginning in 2003, a substantial share of states only report educations in coarse bins. The bins are:

- 8th grade or less
- 9th - 12th grade, no diploma
- High school graduate or GED
- Some college, but no degree
- Associate degree
- Bachelor's degree
- Master's degree
- Doctorate or professional degree
- Unknown

To reduce noise, we slightly coarsen the bins. We put the small share of people people with an 8th grade or less into the HS Dropouts group. We aggregate HS dropouts and middle-school dropouts due to concerns about statistical precision.

We aggregate some college and Associate degrees into the Some College group; it is not clear which group would have higher rank or socioeconomic status on average. To reduce noise, we aggregate Bachelor's, Master's, and Doctorate/professional degrees. These aggregation choices are unlikely to affect our estimates of mortality at the bottom of the distribution.

We exclude deaths of foreign residents.

Missing education. Georgia, Oklahoma, Rhode Island, and South Carolina do not consistently report education data linked to the death certificates. Because their entry and exit from the data could bias mortality trends, we drop mortality records and population totals for these states. The mortality rates we report are thus mortality rates for the remaining states.

The remaining mortality records occasionally report missing education. In each age-gender-race-year category, we obtain the proportion of death records with non-missing educations belonging to each education group. We assign an education group to mortality records where education is

missing, assuming that the missing distribution is the same as the non-missing distribution. For example, if 25% of mortality records with non-missing education in a given cell have a high school degree only, we assign 25% of the mortality records with missing education to have a high school degree. This practice is standard; see Case and Deaton (2015; 2017).

Missing data cannot account for the large mortality changes described in the body of the paper. After dropping the four states, 2.90% of white and 4.34% of black mortality records are missing education (across all years). Roughly 20% of missing deaths among whites (and 25% among blacks) are assigned to high school dropouts; even in the extreme case where *all* of these assignments were incorrect, it would erroneously assign only about 0.6% of white and 1% of black deaths to the bottom education bin, thus creating very little bias.[33]

CPS. Following Case and Deaton (2017), we use the March CPS extracts prepared by the Center for Economic and Policy Research.

Harmonizing educations. A standard issue when harmonizing educational attainment data across several datasets is the treatment of people who drop out of high school in 12th grade. Jaeger (1997) recommends treating 12th grade dropouts as people who complete high school, since in earlier datasets, people who complete part of 12th grade were typically treated as people who completed high-school. Therefore, following Jaeger (1997)'s advice, in the CPS and ACS data which have fine information about educational attainment, we code people who drop out of high school in 12th grade as completing high school.[34]

Early mortality data reports years of school completed. Beginning in 2003, a share of states began reporting education in *categories*, where one of the categories is 9th–12th grade, no diploma. By 2018, all deaths are coded using the 2003 categories. It is not possible to disaggregate the data further. Before 2003, the data only include years of high school (e.g., 1 year of high school or 4 years of high school). For the data coded using the pre-2003 categories, we consider dropouts to be people who attained less than 4 years of high school.

We acknowledge a concern that beginning in 2003, we include people who attained some 12th grade education as dropouts. We emphasize that, due to the death certificate aggregation, there

[33]By 2000 and after, the share of missing data falls by about a third. Thus, the trends over the past 18 years are even less subject to this concern.

[34]This is automatically implemented in CEPR's public-use CPS data. When we augment the CPS data with the ACS data, we recode CEPR's ACS education variable so it is consistent with the CPS data.

is no other way to harmonize the mortality data over time. However, this data limiation is unlikely to play a large role in our results. For example, in the 2018 ACS, there are approximately 5 million people who have 12th grade, no diploma. By contrast, there are 86 million people who do not attain a high-school degree (and have no 12th grade education at all) and 70 million people who attain a high-school diploma. As a result, the share of people who receive some 12th grade education but no degree represents 7% or less of either group.

Moving averages. Because we use small population cells (e.g., white high-school dropouts ages 30–34), to address noise, we use the moving 5-year average for the total population denominator when 5 years of data are available. In 1993 and 2017, we use the moving 3-year average (1992–1994 and 2016–2018, respectively). In 1992 and 2018, we use the predicted values from a regression of population totals on the adjacent years (1992–1994 and 2016–2018, respectively).

- We do not use the 1990 or 1991 CPS data because the education question changed in 1992, so the estimates of the dropouts population is discontinuous in 1991.
- We do not use the 2019 or 2020 CPS because these extracts are not yet harmonized by CEPR.

We pool annual data into 3-year bins to focus on long-term trends and minimize spurious year-on-year variation in results; e.g. for 1992–94, we use average CPS population in each group and the average number of deaths in each group.

Institutionalized populations. The CPS does not survey institutionalized populations, e.g. people living in prisons or hospitals, but deaths in institutions are counted in the mortality records. To obtain accurate mortality rates, we generate institutionalized populations in each year as follows:

1. We obtain counts of the institutionalized population by age, education, gender, race, and year in the 1990 and 2000 U.S. Census (Ruggles et al., 2021) and 2006–2018 American Communities Survey (United States Census Bureau, 2007–2019a).
2. For years between Censuses (1991–1999) or between the Census and American Communities Survey (2001–2005), we impute the number of institutionalized people in each age-education-gender-race-year by generating a linear prediction of the population between nearest surveys. We use the 2006 ACS because it has more accurate counts of the institutionalized populations. For example, if there were 1,000 institutionalized white women aged 50–54 with a high school degree in the 1990 Census and 1,200 in the 2000 Census, we would impute 1,100 in 1995 (where there is no Census available).
3. We add institutionalized populations to our count of the non-institutionalized populations from

the CPS. We compute mortality rates as the number of deaths divided by the total population.

In the ages in our sample, this procedure gives that the share of the white (black) male institutionalized population was 1.3% (6.2%) in 2018. The share of the white (black) female institutionalized population was 0.5% (0.7%). In order for errors from this imputation process to substantially bias mortality estimates, the institutionalized population would need to fluctuate non-linearly in the imputed years. Incarceration rates do rise substantially in the 1990s, but the change is close to linear over time, suggesting that the imputation is a good approximation.

The 2018 CPS does not include people living in college dormitories. We do not adjust for this because we only consider the population older than age 25.

Cause of deaths. We partition all deaths into five groups: cancer, heart disease, deaths of despair, injuries, and other diseases. We construct these groups by using codes from the International Statistical Classification of Diseases and Related Health Problems. NCHS reports ICD-9 codes for 1992–1998 and ICD-10 codes for 1999–2018. We list below the codes pertaining to each cause of death. For consistency, we follow the data appendix and public code from Case and Deaton (2017) to define deaths from cancer, heart disease, and deaths of despair.

- Cancer. ICD-9: 140–208; ICD-10: C (all).
- Heart Disease. ICD-9: 390-429; ICD-10: I0–I9, I11, I13, I20–I51.
- Deaths of Despair. ICD-9: 571, 850–860, 950–959, 980; ICD-10: K70, K73, K74, X40–45, Y10–15, Y45, Y47, Y49, Y87.0.
- Injuries. ICD-9: 800-999 & not a death of despair; ICD-10: V, W, X, Y & not a death of despair.
- Other Diseases. All deaths not otherwise classified.

Table A1 reports the share of deaths among 25–69-year-olds in 2018, ordered by importance. We report the categories used in the paper, and then disaggregate remaining deaths according to major ICD-10 categories.

C Appendix: Methods

C.1 Analytical Proof of NRA Bounds

C.1.1 Proof of Proposition 1

Formalizing the set-up from the text, let $y \sim F$, where F has support $[\underline{y},\overline{y}]$. Put $r_0 = \underline{y}$ and $r_{K+1} = \overline{y}$. In our benchmark case where y denotes survival rates, F has support [0,1]. For simplicity we impose F has finite support in all proofs below. The arguments are identical if F has infinite support except in the top or bottom bin, where the bounds must be adjusted slightly.[35]

Part 1: Find x_k^*. First define $\mathcal{V}_k$ as the set of weakly increasing CEFs which meet the bin mean. Put otherwise, let $\mathcal{V}_k$ be the set of weakly increasing $v : [x_k, x_{k+1}] \to \mathbb{R}$ satisfying

$$r_k = \frac{1}{x_{k+1} - x_k} \int_{x_k}^{x_{k+1}} v(x) dx.$$

Now choose $z \in \mathcal{V}_k$ such that

$$z(x) = \begin{cases} r_{k-1}, & x_k \leq x < j \\ r_{k+1}, & j \leq x \leq x_{k+1}. \end{cases}$$

Note that z and j both exist and are unique (it suffices to show that just j exists and is unique, as then z must be also). We can solve for j by noting that z lies in $\mathcal{V}_k$, so it must meet the bin mean. Hence, by evaluating the integrals, j must satisfy:

$$\begin{aligned} r_k &= \frac{1}{x_{k+1} - x_k} \int_{x_k}^{x_{k+1}} z(x) dx \\ &= \frac{1}{x_{k+1} - x_k} \left(\int_{x_k}^{j} r_{k-1} dx + \int_{j}^{x_{k+1}} r_{k+1} dx \right) \\ &= \frac{1}{x_{k+1} - x_k} ((j - x_k) r_{k-1} + (x_{k+1} - j) r_{k+1}). \end{aligned}$$

Note that these expressions invoke assumption U, as the integration of $z(x)$ does not require any adjustment for the density on the x axis. For a more general proof with an arbitrary distribution of x, see the following section.

[35] We focus on finite-support F to lighten notation. In practice for the analyst, almost every distribution can be restricted with only slight loss of generality to have finite support. For instance, if one studies the CEF of wages given education, $\overline{y}$ can be set to an implausibly high value.

With some algebraic manipulations, we obtain that $j = x_k^*$.

Part 2: Prove the bounds. In the next step, we show that x_k^* is the smallest point at which no $v \in \mathcal{V}_k$ can be r_{k-1}, which means that there must be some larger lower bound on $E(y|x)$ for $x \geq x_k^*$. In other words, we prove that

$$x_k^* = \sup\Big\{x| \text{ there exists } v \in \mathcal{V}_k \text{ such that } v(x) = r_{k-1}.\Big\}.$$

We must show that x_k^* is an upper bound and that it is the least upper bound.

First, x_k^* is an upper bound. Suppose that there exists $j' > x_k^*$ such that for some $w \in \mathcal{V}_k$, $w(j') = r_{k-1}$. Observe that by monotonicity and the bounds from Manski and Tamer (2002), $w(x) = r_{k+1}$ for $x \leq j'$; in other words, if $w(j')$ is the mean of the mean of the prior bin, it can be no lower or higher than the mean of the prior bin up to point j'. But since $j' > j$, this means that

$$\int_{x_k}^{j'} w(x)dx < \int_{x_k}^{j'} z(x)dx,$$

since $z(x) > w(x)$ for all $h \in (j, j')$. But recall that both z and w lie in $\mathcal{V}_k$ and must therefore meet the bin mean; i.e.,

$$\int_{x_k}^{x_{k+1}} w(x)dx = \int_{x_k}^{x_{k+1}} z(x)dx.$$

But then

$$\int_{j'}^{x_{k+1}} w(x)dx > \int_{j'}^{x_{k+1}} z(x)dx.$$

That is impossible by the bounds from Manski and Tamer (2002), since $w(x)$ cannot exceed r_{k+1}, which is precisely the value of $z(x)$ for $x \geq j$.

Second, j is the least upper bound. Fix $j' < j$. From the definition of z, we have shown that for some $h \in (j', j)$, $z(h) = r_{k-1}$ (and $z \in \mathcal{V}_k$). So any point j' less than j would not be a lower bound on the set — there is a point h larger than j' such that $z(h) = r_{k-1}$.

Hence, for all $x < x_k^*$, there exists a function $v \in \mathcal{V}_k$ such that $v(x) = r_{k-1}$; the lower bound on $E(y|x)$ for $x < x_k^*$ is no greater than r_{k-1}. By choosing z' with

$$z'(x) = \begin{cases} r_{k-1}, & x_k \leq x \leq j \\ r_{k+1}, & j < x \leq x_{k+1}, \end{cases}$$

it is also clear that at x_k^*, the lower bound is no larger than r_{k-1} (and this holds in the proposition itself, substituting in x_k^* into the lower bound in the second equation).

Now, fix $x' \in (x_k^*, x_{k+1}]$. Since x_k^* is the supremum, there is no function $v \in \mathcal{V}_k$ such that $v(x') = r_{k-1}$. Thus for $x' > x_k^*$, we seek a sharp lower bound larger than r_{k-1}. Write this lower bound as

$$Y_{x'}^{min} = \min\Big\{ v(x') \text{ for all } v \in \mathcal{V}_k \Big\},$$

where $Y_{x'}^{min}$ is the smallest value attained by any function $v \in \mathcal{V}_k$ at the point x'.

We find this $Y_{x'}^{min}$ by choosing the function which maximizes every point after x', by attaining the value of the subsequent bin. The function which minimizes $v(x')$ must be a horizontal line up to this point.

Pick $\tilde{z} \in \mathcal{V}_k$ such that

$$\tilde{z}(x) = \begin{cases} \underline{Y}, & x_k \leq x' \\ r_{k+1}, & x' < x_{k+1} \end{cases}.$$

By integrating $\tilde{z}(x)$, we claim that $\underline{Y}$ satisfies the following:

$$\frac{1}{x_{k+1} - x_k}((x' - x_k)\underline{Y} + (x_{k+1} - x')r_{k+1}) = r_k.$$

As a result, $\underline{Y}$ from this expression exists and is unique, because we can solve the equation. Note that this integration step also requires that the distribution of x be uniform, and we generalize this argument in the following section.

By similar reasoning as above, there is no $Y' < \underline{Y}$ such that there exists $w \in \mathcal{V}_k$ with $w(x') = Y'$. Otherwise there must be some point $x > x'$ such that $w(x') > r_{k+1}$ in order that w matches the bin means and lies in $\mathcal{V}_k$; the expression for $\underline{Y}$ above maximizes every point after x', leaving no additional room to further depress $\underline{Y}$.

Formally, suppose there exists $w \in \mathcal{V}_k$ such that $w(x') = Y' < \underline{Y}$. Then $w(x') < \tilde{z}(x')$ for all $x < x'$, since w is monotonic. As a result,

$$\int_{x_k}^{x'} \tilde{z}(x)dx > \int_{x_k}^{x'} w(x)dx.$$

But recall that

$$\int_{x_k}^{x_{k+1}} w(x)dx = \int_{x_k}^{x_{k+1}} \tilde{z}(x)dx,$$

so

$$\int_{x'}^{x_{k+1}} w(x)dx > \int_{x'}^{x_{k+1}} \tilde{z}(x)dx.$$

This is impossible, since $\tilde{z}(x) = r_{k+1}$ for all $x > x'$, and by Manski and Tamer (2002), $w(x) \leq r_{k+1}$ for all $w \in \mathcal{V}_k$. Hence there is no such $w \in \mathcal{V}_k$, and therefore $\underline{Y}$ is smallest possible value at x', i.e. $\underline{Y} = Y_{x'}^{min}$.

By algebraic manipulations, the expression for $\underline{Y} = Y_x^{min}$ reduces to

$$Y_x^{min} = \frac{(x_{k+1} - x_k)r_k - (x_{k+1} - x)r_{k+1}}{x - x_k}, \; x \geq x_k^*.$$

The proof for the upper bounds uses the same structure as the proof of the lower bounds.

Finally, the body of this proof gives sharpness of the bounds. For we have introduced a CEF $v \in \mathcal{V}_k$ that obtains the value of the upper and lower bound for any point $x \in [x_k, x_{k+1}]$. For any value y within the bounds, one can generate a CEF $v \in \mathcal{V}_k$ such that $v(x) = y$. □

C.1.2 Analytical Bounds when Uniformity Does Not Hold

Suppose we relax assumption U. We continue to impose for notational simplicity that x is drawn from a continuous distribution, though similar arguments apply to discrete x. We characterize x by some known probability density function, which we assume is integrable in every bin k. Then we derive the following bounds.

Proposition 2. *Let x be in bin k. For simplicity we work with continuous distributions of x. Let $f_k(x)$ be the probability density function of x in bin k. Under assumptions M, I, MI (Manski and Tamer, 2002), and without additional information, the following bounds on $E(y|x)$ are sharp:*

$$\begin{cases} r_{k-1} \leq E(y|x) \leq \frac{r_k - r_{k-1}\int_{x_k}^{x} f_k(s)ds}{\int_{x}^{x_{k+1}} f_k(s)ds}, & x < x_k^* \\ \frac{r_k - r_{k+1}\int_{x}^{x_{k+1}} f_k(s)ds}{\int_{x_k}^{x} f_k(s)ds} \leq E(y|x) \leq r_{k+1}, & x \geq x_k^* \end{cases}$$

where x_k^* satisfies:

$$r_k = r_{k-1}\int_{x_k}^{x_k^*} f_k(s)ds + r_{k+1}\int_{x_k^*}^{x_{k+1}} f_k(s)ds.$$

The proof follows the same argument as in Proposition 1. With an arbitrary distribution, $\mathcal{V}_k$ now constitutes the functions $v:[x_k,x_{k+1}]\to\mathbb{R}$ which satisfy:

$$\int_{x_k}^{x_{k+1}} v(s)f_k(s)ds = r_k.$$

As before, choose $z\in\mathcal{V}_k$ such that

$$z(x) = \begin{cases} r_{k-1}, & x_k \le x < j \\ r_{k+1}, & j \le x \le x_{k+1}. \end{cases}$$

Because the distribution of x is no longer uniform, j must now satisfy

$$\begin{aligned} r_k &= \int_{x_k}^{x_{k+1}} z(s)f_k(s)ds \\ &= r_{k-1}\int_{x_k}^{j} f_k(s)ds + r_{k+1}\int_{j}^{x_{k+1}} f_k(s)ds. \end{aligned}$$

This implies that $j = x_k^*$, precisely.

The rest of the arguments follow identically, except we now claim that for $x > x_k^*$, $\underline{Y} = Y_x^{min}$ satisfies the following:

$$r_k = \int_{x_k}^{x} Y_x^{min} f_k(s)ds + \int_{x}^{x_{k+1}} r_{k+1}f_k(s)ds.$$

By algebraic manipulations, we obtain:

$$Y_x^{min} = \frac{r_k - r_{k+1}\int_x^{x_{k+1}} f_k(s)ds}{\int_{x_k}^{x} f_k(s)ds}$$

and the proof of the lower bounds is complete. As before, the proof for upper bounds follows from identical logic. □

C.1.3 Bounds on μ_a^b

Define

$$\mu_a^b = \frac{1}{b-a}\int_a^b E(y|x)di.$$

Let Y_x^{min} and Y_x^{max} be the lower and upper bounds respectively on $E(y|x)$ given by Proposition 1. We seek to bound μ_a^b when x is observed only in discrete intervals.

Proposition 3. *Let $b \in [x_k, x_{k+1}]$ and $a \in [x_h, x_{h+1}]$ with $a < b$. Let assumptions M, I, MI (Manski and Tamer, 2002) and U hold. Then, if there is no additional information available, the following bounds are sharp:*

$$\begin{cases} Y_b^{min} \leq \mu_a^b \leq Y_a^{max}, & h = k \\ \frac{r_h(x_k - a) + Y_b^{min}(b - x_k)}{b-a} \leq \mu_a^b \leq \frac{Y_a^{max}(x_k - a) + r_k(b - x_k)}{b-a}, & h+1 = k \\ \frac{r_h(x_{h+1} - a) + \sum_{\lambda=h+1}^{k-1} r_\lambda(x_{\lambda+1} - x_\lambda) + Y_b^{min}(b - x_k)}{b-a} \leq \mu_a^b \leq \frac{Y_a^{max}(x_{h+1} - a) + \sum_{\lambda=h+1}^{k-1} r_\lambda(x_{\lambda+1} - x_\lambda) + r_k(b - x_k)}{b-a}, & h+1 < k. \end{cases}$$

The order of the proof is as follows. If a and b lie in the same bin, then μ_a^b is maximized only if the CEF is minimized prior to a. As in the proof of proposition 1, that occurs when the CEF is a horizontal line at Y_x^{min} up to a, and a horizontal line Y_x^{max} at and after a. If a and b lie in separate bins, the value of the integral in bins that are contained between a and b is determined by the observed bin means. The portions of the integral that are not determined are maximized by a similar logic, since they both lie within bins. We prove the bounds for maximizing μ_a^b, but the proof is symmetric for minimizing μ_a^b.

Part 1: Prove the bounds if a and b lie in the same bin. We seek to maximize μ_a^b when $a, b \in [x_k, x_{k+1}]$. This requires finding a candidate CEF $v \in \mathcal{V}_k$ which maximizes $\int_a^b v(x)dx$. Observe that the function $v(x)$ defined as

$$v(x) = \begin{cases} Y_a^{min}, & x_k \leq x < a \\ Y_a^{max}, & a \leq x \leq x_{k+1} \end{cases}$$

has the property that $v \in \mathcal{V}_k$. For if $a \geq x_k^*$, $v = \tilde{z}$ from the second part of the proof of proposition 1. If $a < x_k^*$, the CEF in $\mathcal{V}_k$ which yields Y_a^{max} is precisely v (by a similar argument which delivers the upper bounds in proposition 1).

This CEF maximizes μ_a^b, because there is no $w \in \mathcal{V}_k$ such that

$$\frac{1}{b-a}\int_a^b w(x)dx > \frac{1}{b-a}\int_a^b v(x)dx.$$

Note that for any $w \in \mathcal{V}_k$, $\frac{1}{x_{k+1}-x_k}\int_{x_k}^{x_{k+1}} w(x)dx = \frac{1}{x_{k+1}-x_k}\int_{x_k}^{x_{k+1}} v(x)dx = r_k$. Hence in order that $\int_a^b w(x)dx > \int_a^b v(x)dx$, there are two options. The first option is that

$$\int_{x_k}^{a} w(x)dx < \int_{x_k}^{a} v(x)dx.$$

That is impossible, since there is no room to depress w given the value of v after a. If $a < x_k^*$, then it is clear that there is no w giving a larger μ_a^b, since $r_{k-1} \leq w(x)$ for $x_{k-1} \leq x \leq a$, so w is bounded below by v. If $a \geq x_k^*$, then $v(x) = r_{k+1}$ for all $a \leq x \leq x_{k+1}$. That would leave no room to depress w further; if $\int_{x_k}^{a} w(x)dx < \int_{x_k}^{a} v(x)dx$, then $\int_a^{x_{k+1}} w(x)dx > \int_a^{x_{k+1}} v(x)dx$, which cannot be the case if $v = r_{k+1}$, by the bounds given in Manski and Tamer (2002).

The second option is that

$$\int_{b}^{x_k} w(x)dx < \int_{b}^{x_k} v(x)dx.$$

This is impossible due to monotonicity. For if $\int_a^b w(x)dx > \int_a^b v(x)dx$, then there must be some point $x' \in [a,b)$ such that $w(x') > v(x')$. By monotonicity, $w(x) > v(x)$ for all $x \in [x', x_{k+1}]$ since $v(x) = Y_a^{max}$ in that interval. As a result,

$$\int_{b}^{x_k} w(x)dx > \int_{b}^{x_k} v(x)dx,$$

since $b \in (x', x_{k+1})$. (If $b = x_{k+1}$, then only the first option would allow w to maximize the desired μ_a^b.)

Therefore, there is no such w, and v indeed maximizes the desired integral. Integrating v from a to b, we obtain that the upper bound on μ_a^b is $\frac{1}{b-a}\int_a^b Y_a^{max} dx = Y_a^{max}$. Note that there may be many functions which maximize the integral; we only needed to show that v is one of them.

To prove the lower bound, use an analogous argument.

Part 2: Prove the bounds if a and b do not lie in the same bin. We now generalize the set up and permit $a,b \in [0,100]$. Let $\mathcal{V}$ be the set of weakly increasing functions such that $\frac{1}{x_{k+1}-x_k}\int_{x_k}^{x_{k+1}} v(x)dx = r_k$ for all $k \leq K$. In other words, $\mathcal{V}$ is the set of functions which match the means of every bin. Now

observe that for all $v \in \mathcal{V}$,

$$\begin{aligned}\mu_a^b &= \frac{1}{b-a}\int_a^b v(x)dx \\ &= \frac{1}{b-a}\left(\int_a^{x_{h+1}} v(x)dx + \int_{x_{h+1}}^{x_k} v(x)dx + \int_{x_k}^b v(x)dx\right),\end{aligned}$$

by a simple expansion of the integral.

But for all $v \in \mathcal{V}$,

$$\int_{x_{h+1}}^{x_k} v(x)dx = \sum_{\lambda=h+1}^{k-1} r_\lambda (x_{\lambda+1} - x_\lambda)$$

if $h+1 < k$ and

$$\int_{x_{h+1}}^{x_k} v(x)dx = 0$$

if $h+1 = k$. For in bins completely contained inside $[a,b]$, there is no room for any function in $\mathcal{V}$ to vary; they all must meet the bin means.

We proceed to prove the upper bound. We split this into two portions: we wish to maximize $\int_a^{x_{h+1}} v(x)dx$ and we also wish to maximize $\int_{x_k}^b v(x)dx$. The values of these objects are not codependent. But observe that the CEFs $v \in \mathcal{V}_k$ which yield upper bounds on these integrals are the very same functions which yield upper bounds on $\mu_a^{x_{h+1}}$ and $\mu_{x_k}^b$, since $\mu_s^t = \frac{1}{t-s}\int_s^t v(x)dx$ for any s and t. Also notice that a and x_{h+1} both lie in bin h, while b and x_k both lie in bin k, so we can make use of the first portion of this proof.

In part 1, we showed that the function $v \in \mathcal{V}$, $v : [x_h, x_{h+1}] \to \mathbb{R}$, which maximizes $\mu_a^{x_{h+1}}$ is

$$v(x) = \begin{cases} Y_a^{min}, & x_h \leq x < a \\ Y_a^{max}, & a \leq x \leq x_{h+1}. \end{cases}$$

As a result

$$\max_{v \in \mathcal{V}} \left\{ \int_a^{x_{h+1}} v(x)dx \right\} = \int_a^{x_{h+1}} Y_a^{max} dx = Y_a^{max}(x_{h+1} - a).$$

Similarly, observe that x_k and b lie in the same bin, so the function $v : [x_k, x_{k+1}] \to \mathbb{R}$, with $v \in \mathcal{V}$

which maximizes $\int_{x_k}^{b} v(x)dx$ must be of the form

$$v(x)=\begin{cases} Y_{x_k}^{min}, & x_k \leq x < a \\ Y_{x_k}^{max}, & b \leq x \leq x_{k+1}. \end{cases}$$

With identical logic,

$$\max_{v\in\mathcal{V}}\left\{\int_{x_k}^{b} v(x)dx\right\}=\int_{x_k}^{b} Y_{x_k}^{max}dx=Y_{x_k}^{max}(b-x_k).$$

And by proposition 1, $x_k \leq x_k^*$ so $Y_{x_k}^{max}=r_k$. (Note that if $x_k=x_k^*$, substituting x_k^* into the second expression of proposition 1 still yields that $Y_{x_k}^{max}=r_k$.)

Now we put all these portions together. First let $h+1=k$. Then $\int_{x_{h+1}}^{x_k} v(x)dx=0$, so we maximize μ_a^b by

$$\frac{1}{b-a}(Y_a^{max}(x_{h+1}-a)+r_k(b-x_k)).$$

Similarly, if $h+1<k$ and there are entire bins completely contained in $[a,b]$, then we maximize μ_a^b by

$$\frac{1}{b-a}\left(Y_a^{max}(x_{h+1}-a)+\sum_{\lambda=h+1}^{k-1} r_\lambda(x_{\lambda+1}-x_\lambda)+r_k(b-x_k)\right).$$

The lower bound is proved analogously. Sharpness is immediate, since we have shown that the CEF which delivers the endpoints of the bounds lies in $\mathcal{V}$. As a result, there is a function delivering any intermediate value for the bounds. □

C.2 Numerical Calculation of NRA Bounds with Arbitrary Structural Assumptions

This section describes the numerical optimization approach for calculating bounds on mortality within arbitrary percentiles of the education distribution. To calculate bounds, we discretize the mortality-education relationship and solve a numerical optimization problem that obtains the highest and lowest possible values of expected mortality at any given point, that are consistent with matching the empirical data points and meeting a set of structural constraints on the functional form of the CEF. We first formalize the computational procedure and then describe its numerical implementation.

C.2.1 Conceptual Approach: Functions of the CEF

We write the conditional expectation function in the form $Y(x) = s(x,\gamma)$, where γ is a finite-dimensional vector that lies in parameter space G and serves to parameterize the CEF through the function s. For example, we could estimate the parameters of a linear approximation to the CEF by defining $s(x,\gamma) = \gamma_0 + \gamma_1 * x$. We can approximate an arbitrary nonparametric CEF by defining γ as a vector of discrete values that give the value of the CEF in each of N partitions; we take this approach in our numerical optimizations, setting N to 100.[36] Any statistic m that is a single-valued function of the CEF, such as the average value of the CEF in an interval (μ_a^b), or the slope of the best fit line to the CEF, can be defined as $m(\gamma) = M(s(x,\gamma))$.

Let $f(x)$ represent the probability distribution of x. Define Γ as the set of parameterizations of the CEF that obey monotonicity and minimize mean squared error with respect to the observed interval data:

$$\Gamma = \underset{g \in G}{\operatorname{argmin}} \sum_{k=1}^{K} \left\{ \int_{x_k}^{x_{k+1}} f(x)dx \left(\left(\frac{1}{\int_{x_k}^{x_{k+1}} f(x)dx} \int_{x_k}^{x_{k+1}} s(x,g) f(x)dx \right) - \overline{r}_k \right)^2 \right\} \quad \text{(C.1)}$$

such that

$s(x,g)$ is weakly increasing in x. (Monotonicity)

Decomposing this expression, $\frac{1}{\int_{x_k}^{x_{k+1}} f(x)dx} \int_{x_k}^{x_{k+1}} s(x,g) f(x)dx$ is the mean value of $s(x,g)$ in bin k, and $\int_{x_k}^{x_{k+1}} f(x)dx$ is the mass in bin k. The minimand is thus a bin-weighted MSE.[37] Recall that for the rank distribution, $x_1 = 0$ and $x_{K+1} = 100$. We can easily add additional structural constraints on $s(x,\gamma)$ to Equation C.1 (e.g., the curvature constraint below) or remove the monotonicity constraint.

The bounds on $m(\gamma)$ are therefore:

$$\begin{aligned} m^{min} &= \inf\{m(\gamma) \mid \gamma \in \Gamma\} \\ m^{max} &= \sup\{m(\gamma) \mid \gamma \in \Gamma\}. \end{aligned} \quad \text{(C.2)}$$

For example, bounds on the best linear approximation to the CEF can be defined by the following process. First, consider the set of all CEFs that satisfy monotonicity and minimize mean-squared error

[36] For example, $s(x,\gamma_{50})$ would represent $E(y|x \in [49,50])$.

[37] While we choose to use a weighted mean squared error penalty, in principle Γ could use other penalties.

with respect to the observed bin means.[38] Next, compute the slope of the best linear approximation to each CEF. The largest and smallest slope constitute m^{min} and m^{max}. Note that this definition of the best linear approximator to the CEF corresponds to the *least squares set* defined by Ponomareva and Tamer (2011).

The set of CEFs that describe the upper and lower bounds in Proposition 1 are step functions with substantial discontinuities. If such functions are implausible descriptions of the data, then the researcher may wish to impose an additional constraint on the curvature of the CEF, which will generate tighter bounds. For example, examination of the mortality-income relationship (which can be estimated at each of 100 income ranks, displayed in Appendix Figure C1) suggests no such discontinuities. Alternately, in a context where continuity has a strong theoretical underpinning but monotonicity does not, a curvature constraint can substitute for a monotonicity constraint and in many cases deliver useful bounds.

We consider a curvature restriction with the following structure:

$$s(x,\gamma) \text{ is twice-differentiable and } |s''(x,\gamma)| \leq \overline{C}. \qquad \text{(Curvature Constraint)}$$

This is analogous to imposing that the first derivative is Lipshitz.[39] Depending on the value of $\overline{C}$, this constraint may or may not bind.

The most restrictive curvature constraint, $\overline{C}=0$, is analogous to the assumption that the CEF is linear. Note that the default practice in many studies of mortality is to estimate the best linear approximation to the CEF of mortality given education (e.g., Cutler et al. (2011) and Goldring et al. (2016)). A moderate curvature constraint is therefore a *less* restrictive assumption than the approach taken in many studies. In practice, we slightly adjust the curvature constraint condition by generating a "normalized" curvature constraint which imposes that the absolute value of the second derivative, divided by the mean mortality across all percentiles, does not exceed a certain $\overline{C}$. The advantage of this approach is that it more readily permits using a conservative value of

[38] In many cases, and in all of our applications, there will exist many such CEFs that exactly match the observed data and the minimum mean-squared error will be zero.

[39] Let X,Y be metric spaces with metrics d_X,d_Y respectively. The function $f: X \to Y$ is *Lipschitz continuous* if there exists $K \geq 0$ such that for all $x_1,x_2 \in X$,

$$d_Y(f(x_1),f(x_2)) \leq K d_X(x_1,x_2).$$

$\overline{C}$ across all groups. We discuss the choice of curvature restriction below.

C.2.2 Computational Approach

This section describes a method to numerically solve the constrained optimization problem suggested by Equations C.1 and C.2.

To make the problem numerically tractable, we solve the discrete problem of identifying the feasible mean value taken by $E(y|x)$ in each of N discrete partitions of x. We thus assume $E(y|x)=s(x,\gamma)$, where γ is a vector that defines the mean value of the CEF in each of the N partitions. We use $N=100$ in our analysis, corresponding to integer rank bins, but other values may be useful depending on the application. In other words, we will numerically calculate upper and lower bounds on $E(y|x\in[0,1])$, $E(y|x\in[1,2])$, ..., $E(y|x\in[99,100])$. Given continuity in the latent function, the discretized CEF will be a very close approximation of the continuous CEF; in our applications, increasing the value of N increases computation time but does not change any of our results.

We solve the problem through a two-step process. Define a N-valued vector $\hat{\gamma}$ as a candidate CEF. First, we calculate the minimum MSE from the constrained optimization problem given by Equation C.1. Put another way, each set Γ is associated to a minimmum MSE (the value of the objective), which we denote $\underline{\text{MSE}}$. We then run a second pair of constrained optimization problems that respectively minimize and maximize the value of $m(\hat{\gamma})$, with the additional constraint that the MSE is equal to the value obtained in the first step, denoted $\underline{\text{MSE}}$. Equation C.3 shows the second stage setup to calculate the lower bound on $m(\hat{\gamma})$:

$$m^{min} = \min_{\hat{\gamma}\in[0,100]^N} m(\hat{\gamma}) \tag{C.3}$$

$$\text{such that}$$

$$s(x,\hat{\gamma}) \text{ is weakly increasing in } x \qquad \text{(Monotonicity)}$$

$$|s''(x,\hat{\gamma})| \leq \overline{C} \qquad \text{(Curvature)}$$

$$\sum_{k=1}^{K}\left[\frac{\|X_k\|}{100}\left(\left(\frac{1}{\|X_k\|}\sum_{x\in X_k} s(x,\hat{\gamma})\right)-\overline{r}_k\right)^2\right] = \underline{\text{MSE}} \qquad \text{(MSE Minimization)}$$

X_k is the set of discrete values of x between x_k and x_{k+1} and $\|X_k\|$ is the width of bin k. $\overline{r}_k$ is the observed mortality in education bin k, and $\underline{\text{MSE}}$ is the lowest mean-squared error obtainable

out of the entire set of education-mortality functions, which is typically zero. The complementary maximization problem obtains the upper bound on $m(\hat{\gamma})$. Note that this particular setup is specific to the uniform rank distribution, but setups with other distributions would be similar.

The purpose of the two-step process is that it is difficult to numerically solve for every possible member of Γ under arbitrary constraints. We render the problem tractable by recasting the problem as a minimization problem of obtaining MSE in the first step. Then, in the second step, we use MSE as a constraint.

Note that setting $m(\gamma)=\gamma_x$ (the x^{th} element of γ) obtains bounds on the value of the CEF at point x. Calculating this for all ranks x from 1 to 100 generates analogous bounds to those derived in proposition 1, but satisfying the additional curvature constraint. Similarly $m(\gamma)=\frac{1}{b-a}\sum_{x=a}^{b}\gamma_x$ yields bounds on μ_a^b.

The numerical method can easily permit the curvature constraint to vary over the CEF. For example, one might believe that there are discontinuities in the CEF at bin boundaries, due to sheepskin effects (Hungerford and Solon, 1987); high-school graduates, upon receiving a diploma, could experience discretely lower mortality probability due to better labor-market outcomes. In other settings, researchers might impose that the CEF has a large (but finite) curvature in one portion of its domain and be more constrained elsewhere.

Calibrating a curvature constraint. This subsection explains how we obtain a benchmark for the curvature constraint using data from Chetty et al. (2016) on mortality rates for U.S. men and women above age 40 from years 2001–2014. We collapse the data to three-year periods and five-year age groups.[40,41] We then use OLS to fit fifth-order polynomials to the mortality-income percentile data we observe. We show two examples of these best-fit functions for 50–54 year old men and women in 2012–2014 (the last year data are available), as well as the range for the normalized second derivative for these groups, in Appendix Figure C1. The $\overline{C}$ we use is 50% larger than the maximum absolute value of the union of these ranges across all such groups we observe.

We construct a $\overline{C}$ that holds for all mortality-income functions as follows. Using the estimated polynomial fit, we analytically compute the absolute value of the second derivative of the best-fit polynomial at every value for every polynomial function. To generate comparable $\overline{C}$, we construct

[40]Since the number of years is not divisible by 3, we group years 2001 and 2002.

[41]Because people are ranked within the percentile for their own age, gender, and year, this departs slightly from the ranking procedure we use in the text.

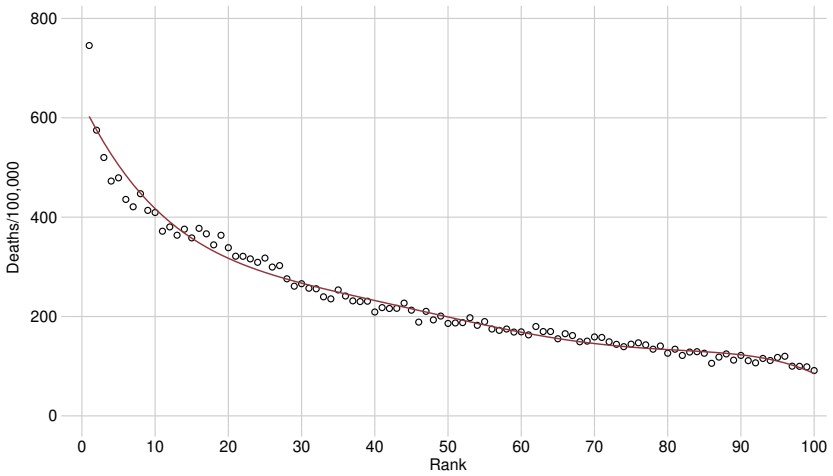
a normalized $\overline{C}$ that accounts for differences in mortality levels by dividing the absolute value of the second derivative by the mean mortality in that age-year group (across all percentiles). Expressed in these terms (and multiplied by 100), the normalized $\overline{C}$ represents the absolute value of the second derivative as a percent of the mean.

Across all age and years, the largest $\overline{C}$ is approximately 2.0%. Because polynomial fits can be inaccurate near the tails, we also compute the largest second-derivative (normalized by the mean mortality across all percentiles) within the set of percentiles [5,95]. This value is 1.5%. We choose a conservative curvature constraint of 3% — about 50% larger than the largest normalized $\overline{C}$ observed in the data.

We acknowledge the concern that mortality-income CEFs may exhibit different curvature than mortality-education CEFs. We argue that the mortality-income CEF at least provides a natural benchmark for how curved the CEF might be in the education setting; we are not aware of another continuous conditioning value to calibrate $\overline{C}$. Moreover, we are comforted that our results are robust to relaxing the curvature constraint altogether.

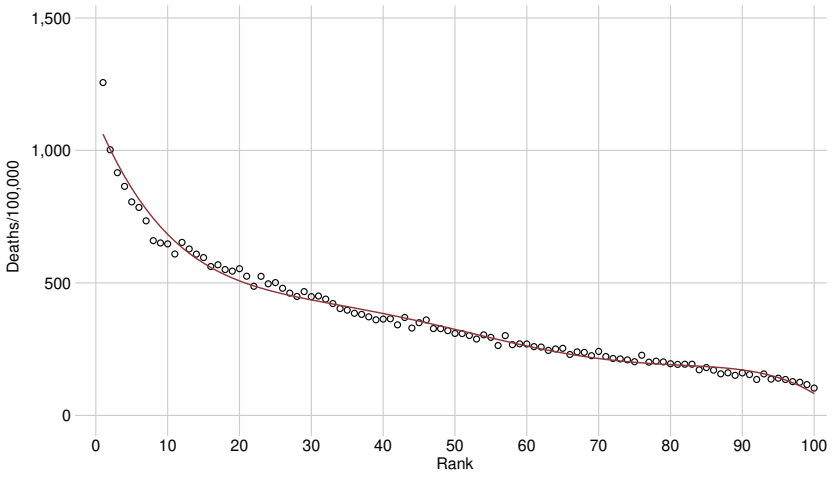
Figure C1
Fifth-Order Polynomial Approximations to the Empirical Mortality-Income CEF

Panel A: 50–54 Year-Old Women in 2012–2014



2nd derivative in [–0.349%,0.876%]
2nd derivative in range [5, 95] in [–0.210%,0.683%]

Panel B: 50–54 Year-Old Men in 2012–2014



2nd derivative in [–0.562%,1.310%]
2nd derivative in range [5, 95] in [–0.339%,1.008%]

Figure C1 presents estimates of the conditional expectation function of U.S. mortality given income rank, using data from Chetty et al. (2016). The CEF is fitted using a fifth-order polynomial. The function plots the best polynomial fit to the data series, and the circles plot the underlying data. The text under the graph shows the range of the second derivative, divided by the mean mortality rate in all percentiles, across the function's support.

C.3 Comparison with Other Approaches

The selection bias in estimates of mortality change among the less educated is widely recognized and has been examined in other studies. Meara et al. (2008), Bound et al. (2015), Hendi (2015), and Leive and Ruhm (2020) (henceforth MBHL) adjust for this bias by randomly reassigning deaths to different education bins, so that bin sizes are comparable across time. For example, to obtain an estimate of mortality in the bottom quartile of the education distribution in 1992 when 19% of 50-year-old men are dropouts and an additional 36% are high school graduates, they would reassign $6/36 = 16.6\%$ of the high school graduate population and 16.6% of high school graduate deaths to the bottom bin.

This approach is equivalent to assuming that the conditional expectation function of mortality given latent education rank takes a specific functional form—a step function that is totally flat in each education category, and has a discrete jump at each education boundary. To be concrete, under this assumption, an individual who just barely managed to complete high school (and thus has the lowest latent education rank among high school graduates) has exactly the same implied socioeconomic status and expected mortality risk as a high school graduate who was right at the margin of completing a two year college degree (and thus has the highest latent education rank among high school graduates). Standard human capital theory suggests that the true functional form is not flat in each category—the high school educated individuals who were at the margin of completing some higher education would have had higher socioeconomic status than those who barely made it to high school, and thus lower mortality risk.

This implicit functional form is nevertheless considered a valid functional form in our bounding exercise, which allows arbitrary steps and slope changes at education boundaries. But this functional form underestimates mortality among the least educated in all periods, because it constructs bins by combining dropouts with average high school graduates – even though the high school graduates with the lowest latent ranks are likely to have higher mortality than average high school graduates. The downward bias on mortality among the least educated will be the highest when the education-mortality gradient is steep. Because this gradient has steepened over time (Goldring et al., 2016), the downward bias on mortality is higher in 2018 than in 1992, which means that mortality change among the least educated is biased downward when we use this functional form.

Note finally that these other approaches are likely to be increasingly biased when the bin boundary

shifts more over time, or when the desired outcome percentiles are very different from the bin boundaries in the raw data, but the estimates from the MBHL function will not reflect this source of error. One reason that few of these authors focus on the very bottom of the education distribution (or the percentiles approximating high school dropouts) is that the large population change in dropouts (among women, from 19% of the population in 1992 to 8% in 2018) leads to a substantial potential for bias. With our approach, in contrast, the bounds reflect the uncertainty in the estimates and become wider in cases like these where bin boundaries have shifted substantially. Our bounds thus accurately convey the uncertainty due to misalignment between desired outcome percentiles and bin boundaries in the data.

The MBHL function generates a mortality estimate among the least educated that is close to our lower bound mortality estimate. Our results are therefore entirely consistent with Meara et al. (2008), who find that death rates among those with a high school education or less are diverging from those with any college education. We find similar effects for the period up to 2000 studied by Meara et al. (2008), and show that (i) mortality by education continues to diverge from 2000–2018; and (ii) the bottom 10% of whites do particularly badly in both periods.

In contrast, Bound et al. (2015) argue that the composition adjustment effectively erases large mortality increases among non-Hispanic whites in the bottom 25%, though they continue to find average decreases in life expectancy at age 25 among non-Hispanic whites in this education group. These differences can be reconciled with our finding of substantially rising mortality rates among the least educated non-Hispanic whites. First, as noted in this section, the Bound et al. estimates are at the lower bound of mortality change because of the implicit functional form assumption. Second, we find that mortality increases are most severe among the bottom *10%*; extending the interval to the bottom 25% substantially attenuates the estimated mortality change, and makes the functional form bias larger. Third, we focus on middle-age mortality change, because of known problems of age inaccuracy among older ages (Olshansky et al., 2012); trends among individuals aged 70 and older may substantially influence life expectancy and be different from those studied here.

Hendi (2015, 2017) use data from NHIS to argue that mortality is rising for white women without high school education but not white men. We find worse outcomes for both women and men because: (i) the random reassignment approach of Hendi (2015, 2017) biases downward estimates of mortality change; (ii) as noted in detail by Sasson (2017), the NHIS consistently samples a healthier population than that reflected in population vital statistics, and the mortality followup sample sizes are too

small to precisely estimate mortality changes for less educated groups. We perform a similar analysis in Appendix D.6, showing that while NHIS generates point estimates of lower mortality changes for less educated women (but not men), the estimates are extremely noisy and our estimates are well within the 95% confidence intervals of the NHIS measures. In some age/education groups, NHIS mortality changes are estimated from just a handful of deaths.

Hendi (2015, 2017) also raise the possibility that over-reporting of education in death statistics may have changed over time, as noted by Sorlie and Johnson (1996), leading to underestimates of death rates among older cohorts of dropouts. While we cannot rule out this form of bias, we present a range of evidence in Appendix D (summarized in Section 4.5) suggesting that misreporting of education in death records cannot explain larges increases in mortality among white men and women.

D Appendix: Robustness

D.1 Loosening the Monotonicity Assumption

The results in the body of the paper use an assumption that mortality is weakly monotonically decreasing in the latent education rank. In this section, we explore the sensitivity of the results to loosening this assumption. Using the numerical optimization described in Section C.2, we alter the monotonicity assumption such that the discrete CEF is permitted to be non-monotonic across at most m rank bins out of 100.

Table D1 shows how the bounds on white male and female mortality change in percentiles 0–10 under values of m ranging from 0 to 100. The case of $m = 0$ corresponds to the monotonicity assumption used in the body of the paper and reproduces the results from the main analysis (Figures 6A and 6B in the body of the paper).

Note first that for the 2016–18 results, loosening monotonicity has very little effect. In this period, exactly 10% of men are high school dropouts, meaning μ_0^{10} is known. Among women, 8.0% are dropouts, which gives us μ_0^8; this leaves little room for mortality to depend on functional form assumptions in the bottom 10%.

In 1992–94, dropouts represent the bottom 17.1% and 17.4% for men and women respectively, leaving more uncertainty over the value of mortality in the bottom 10%. Loosening the monotonicity restrictions in 1992 naturally widens the bounds. Figure D1 shows the pair of optimized CEFs that produce the upper and lower bounds for the case of 50–54-year-old white women.

The primary results of rising mortality among both groups are upheld in all cases. The bounds expand moderately up to $m = 20$, rising from [+105%,+152%] for white women at $m = 0$ to [+67%,+204%] at $m = 20$. As conveyed by the figure, loosening monotonicity entirely results in very implausible functional forms for the CEF. Even the CEFs at $m = 20$ are difficult to reconcile with an a priori theory of mortality change, and exhibit higher degrees of non-monotonicity than suggested by the data for any of the subgroups. We therefore view this exercise largely as a demonstration of our method's capability to be adapted to any set of structural assumptions.

Note that if monotonicity and curvature are both totally unrestricted, then much wider bounds are possible, though again we would not consider these to be realistic. For instance, mortality among white men in 1992–94 is known to be 1197 per 100,000 in the bottom 17%. The mathematical

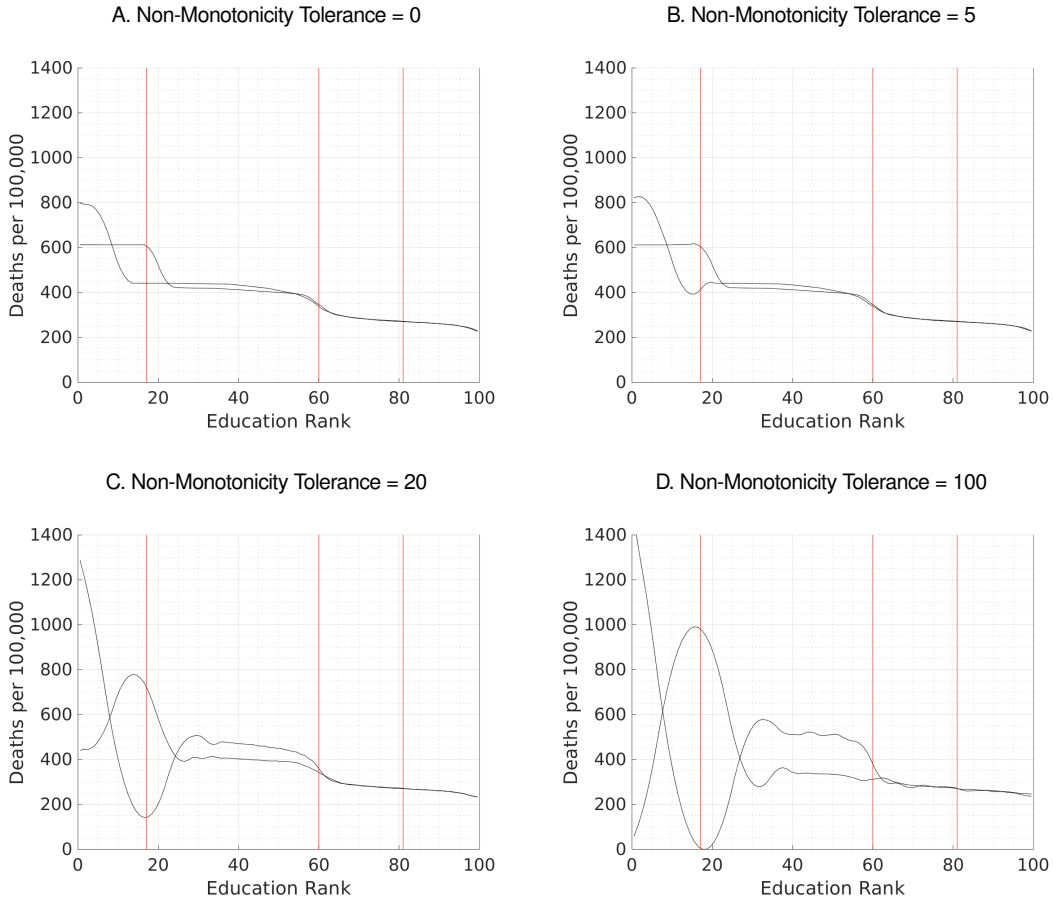
upper bound on mortality in the bottom 10%, with no structural restrictions, would be given by a mortality rate of 2035/100,000 for percentiles 0–10, and 0/100,000 for percentiles 10–17. While mathematically possible, this function is not a realistic possibility. In our setup, it violates both the monotonicity constraint (since mortality would need to rise with high school completion) and the curvature constraint (since implied mortality falls sharply and discontinuously between 10 and 11).

Table D1
Mortality Change in the Least Educated 10%
Estimates with Variable Non-Monotonicity

Group	Non-monotonicity Tolerance	Mortality 1992–94	Mortality 2016–18	Percent Change
White Women, Age 50	0	[613,722]	[1481,1542]	[105.2%,151.6%]
	5	[612,742]	[1477,1546]	[99.2%,152.6%]
	10	[577,753]	[1477,1553]	[96.1%,169.0%]
	15	[524,821]	[1477,1552]	[79.9%,196.1%]
	20	[513,885]	[1475,1556]	[66.6%,203.5%]
	100	[384,945]	[1475,1562]	[56.0%,306.6%]
White Men, Age 50	0	[1200,1394]	[1946,1946]	[39.6%,62.2%]
	5	[1200,1427]	[1946,1946]	[36.4%,62.2%]
	10	[1182,1528]	[1946,1946]	[27.4%,64.7%]
	15	[1161,1590]	[1946,1946]	[22.5%,67.7%]
	20	[1101,1635]	[1946,1946]	[19.0%,76.9%]
	100	[874,1713]	[1946,1946]	[13.6%,122.7%]

The table shows bounds on mortality and mortality change of white men and women aged 50–54 in the least educated 10% of the own-gender education distribution. Bounds are calculated under different tolerances for non-monotonic CEFs. The tolerance column indicates the number of rank cells out of 100 where mortality is permitted to be increasing with higher education.

Figure D1
Bounds on Mortality in the Bottom 10%:
Mortality CEFs under Variable Non-Monotonicity Constraints



The figure shows the conditional expectation functions that generate the highest and lowest values of mortality among the least educated 10%, for 50–54-year-old white women, under different monotonicity assumptions. The tolerance value is the number of rank cells out of 100 where mortality is permitted to be increasing with higher education.

D.2 Robustness to Alternate Specifications and Assumptions

This section reports changes in mortality from 1992–1994 to 2016–2018 calculated under different assumptions and parameters. We focus on the sensitivity of estimates in Figure 6 for non-Hispanic white women; results for other groups are similarly robust to the specifications here.

Percentile Bins Defined in 1992. The main figure was calculated using education percentile bins from approximately the middle of the sample in 2003. In that year dropouts accounted for percentiles 0–10, high school graduates for percentiles 10–45, individuals with some college for percentiles

45–70, and individuals with a B.A. or higher accounted for the top 30%. In 1992, these four education levels respectively represented percentiles 0–17, 17–60, 60–81 and 81–100. Panel A of Figure D2 shows estimates of mortality change from 1992–94 to 2016–2018 using the latter bin boundaries. The broad patterns of mortality change are the same. Mortality changes are slightly smaller in the bottom group, but this is what we would expect given that the bottom group is now defined as the bottom 17% rather than the bottom 10%. The overall divergence of mortality by education is unambiguous.

Ranking within Race and Gender. The body of the paper ranks individuals against members of their own gender. It thus reports, for example, changes in mortality for white women in the bottom 10% of the female education distribution. An alternate approach is to define percentiles within race and gender, and thus to examine mortality changes for white women in the bottom 10% of the white women's education distribution. This approach would be sensible if one's relative position in the own-race socioeconomic distribution was an important factor for mortality. Panel B of Figure D2 presents the main estimates, where education percentiles are defined within race and gender. The pattern of dramatically rising mortality for whites in the bottom 10% remains evident. The total increases in mortality are slightly less under this definition, because white education has increased more than black education over this period, but the change in selection bias is small.

Alternate Bounding Assumptions. The main analysis calculated bounds under the assumptions that (i) expected mortality is weakly monotonically declining in latent educational rank for all groups; and (ii) there is an upper bound on the size of kinks or discrete jumps in the expected mortality function. While these assumptions are sensible and consistent with other research and data on the expectation of mortality as a function of education, we can readily relax these assumptions. Although our main estimates are necessarily less precise under more general assumptions, none any of the substantive conclusions change.

Appendix D.1 above presents bounds on mortality change as we gradually loosen the monotonicity assumption, and allow mortality to rise with education.

Panel C of Figure D2 presents results when we allow the mortality function to have discrete jumps or kinks of any size when crossing educational boundaries, for example, when crossing the threshold of high school completion. Thus Panel C addresses concerns about sheepskin effects by permitting (but

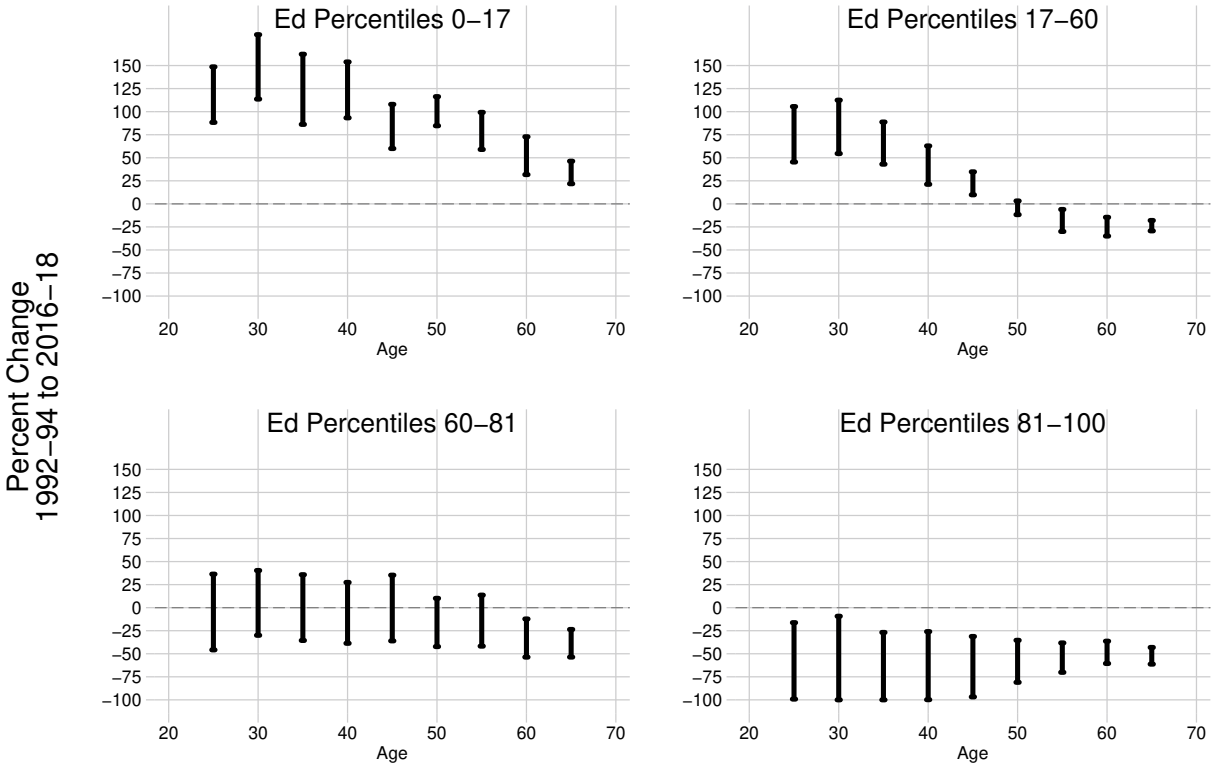
not imposing) mortality to fall discontinuously at the margin of completing a given education level.

Panel D removes the curvature restriction entirely; in this specification, the CEF can have discrete kinks or jumps at any point in the education rank distribution, though it must remain monotonic. In both cases, the key result of substantial mortality gain among the least educated is robust. The other findings of the paper—education divergence among black men and women, and the patterns of deaths of despair—are similarly robust to these specifications.

Figure D2

Change in non-Hispanic White Female Mortality: Sensitivity Analysis

Panel A: 1992 Percentile Boundaries



Panel B: Within Race-Gender Percentiles

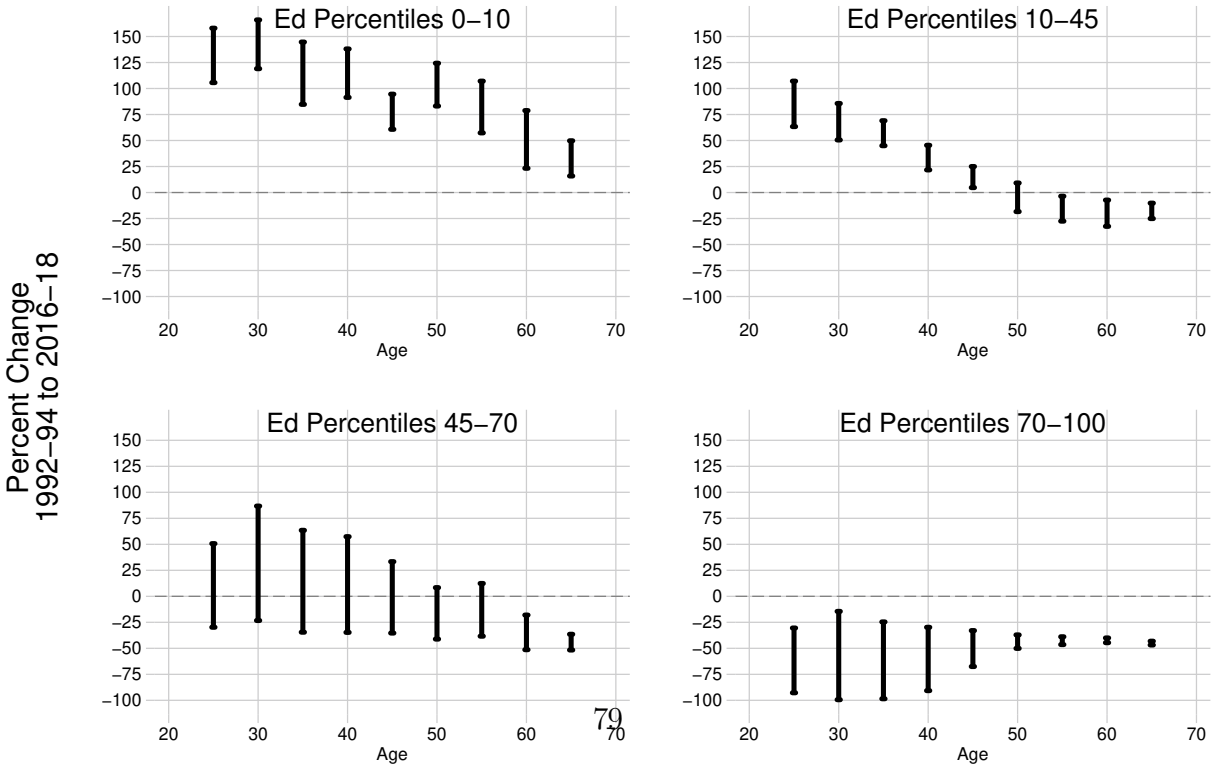
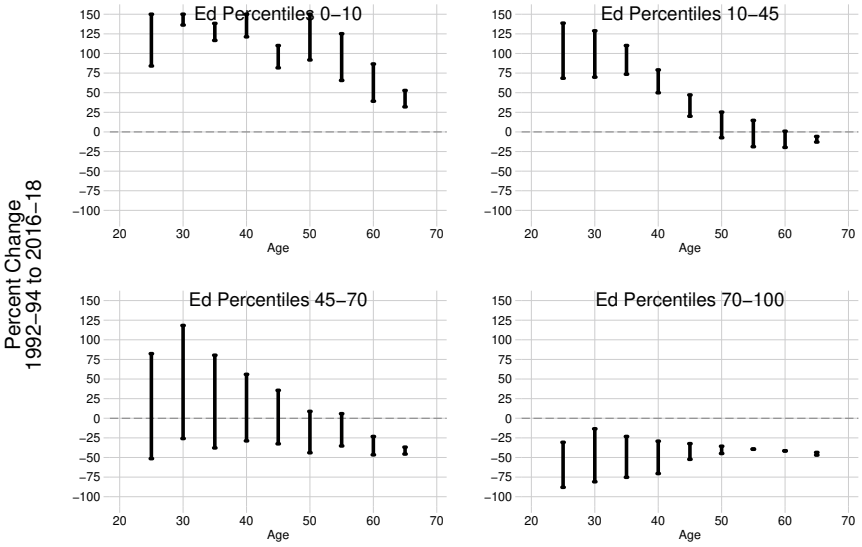


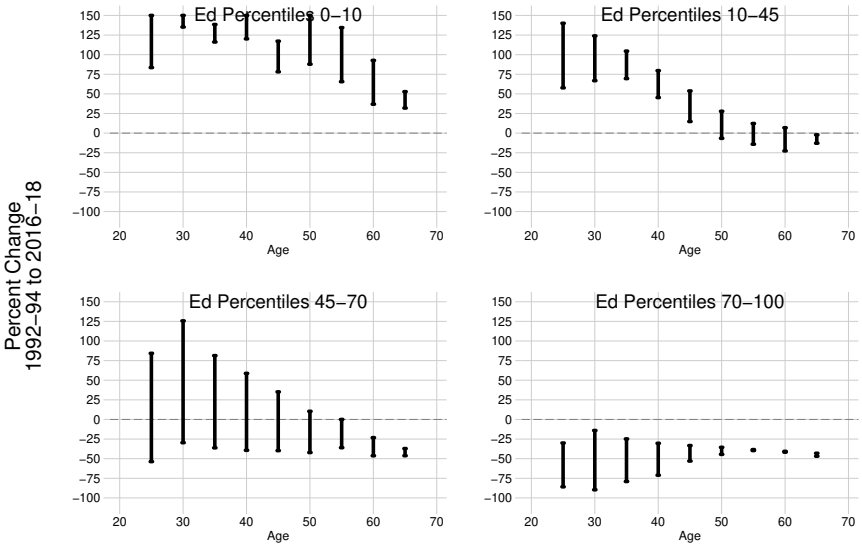
Figure D2

Change in non-Hispanic White Female Mortality: Sensitivity Analysis (Continued)

Panel C: Allow Sheepskin Effects



Panel D: No Curvature Constraint



Note: The figure shows bounds on mortality change from 1992–1994 to 2016–2018, for non-Hispanic white women, by age and percentile education bin, under alternate assumptions from the main body of the paper. The figure is analogous to Panel A of Figure 6, but with different bounding assumptions. Panel A defines education bins boundaries according to education levels in 1992–1994. Panel B defines an individual's education percentile according to the individual's rank in the own-race and own-gender education distribution, rather than in the own-gender education distribution. Panel C allows sheepskin effects, by allowing kinks and discrete jumps at education boundaries (e.g. the rank separating dropouts from high school completers). Panel D estimates bounds on mortality without restricting the curvature of the mortality-education CEF. The lines show the bounded set containing the percentage change in the mortality rate from 1992–1994 to 2016–2018 for the given group.

D.3 Changing Racial Composition

This section examines the hypothesis that relative socioeconomic status *within* education bins has changed for blacks relative to whites during the study period. This kind of change could bias our estimates of mortality change, because we assume that the latent education ranks of blacks and whites *within* the bottom 10% (and within other percentile groups) have not changed during the study period. For example, if whites within the bottom 10% were clustered at the top of this percentile bin in 1992–94 and at the bottom of this bin in 2016–18, then we would expect their measured mortality to rise even if the underlying mortality-education-rank CEF is unchanged. To be concrete, suppose the mean white woman, conditional on being in the bottom 10% of all women, moved from the 7th percentile to the 3rd percentile. Then comparing average mortality among white women in the bottom 10% could still be subject to selection bias, even in our constant composition estimates, because the average white woman in the bottom 10% would be more negatively selected over time.

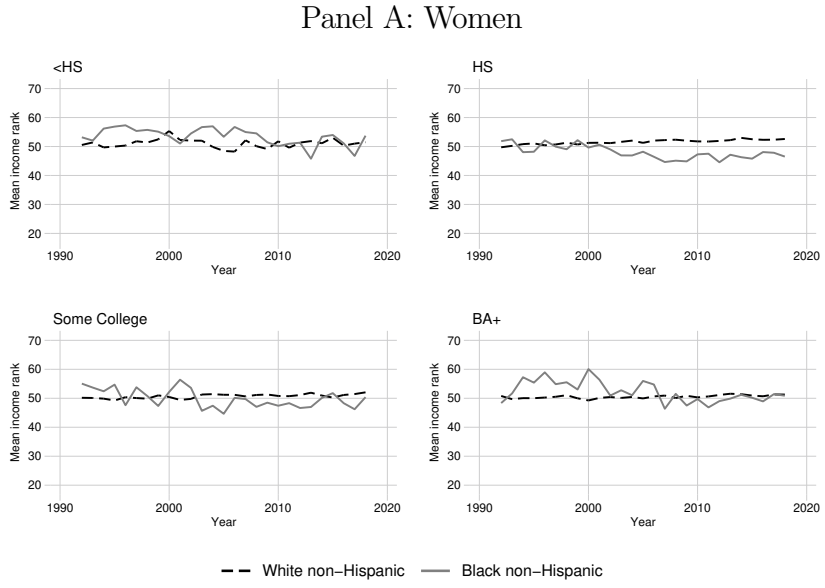
Note that Panel B of Figure D2 already rules this out as a primary explanation for rising white mortality, by showing that mortality is rising for the bottom 10% of whites, not just for whites in the bottom 10% of the national education distribution. Nevertheless, in this section we explore the possibility that the relative status of whites in the bottom 10% of the entire educational distribution has shifted relative to blacks.

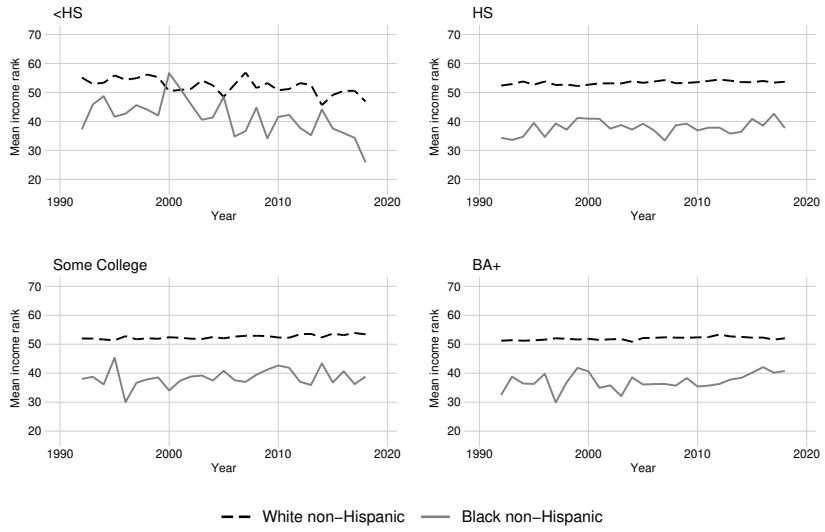
Because education data is interval censored (i.e. we only know that someone is a high school dropout, but we do not observe how close they were to completing high school), we cannot measure the education percentile more precisely. However, we can examine whether the socioeconomic status of white dropouts has changed relative to black dropouts on other measures. We focus on income as measured in the Current Population Survey.

First, we use the granular information on incomes in the CPS to rank all people by income within each gender, age and education bin in each year. We then compute the mean income rank for whites and blacks within each of the four education groups used in the body of the paper. Figure D3 plots the results of this exercise for women and men aged 50–54. The figures show that mean income ranks for blacks and whites, conditional on education level, have remained stable over time. Among women, black and white dropouts have approximately equal income rank throughout the sample period. Among high school graduates, the relative status of whites is increasing relative to that of

blacks, which would bias us *against* finding increases in mortality. Among men, whites have higher average income ranks within every education bin, but their relative advantage is stable over time. Changing relative latent education rank *within* observed education bins therefore cannot explain any of the rise in mortality of white dropouts.

Figure D3
Average Income Ranks for 50–54 year old whites and blacks





Note: The figure shows the average income rank within age, gender and education bins for non-Hispanic white and non-Hispanic black men and women at different education levels. Source: CPS.

D.4 Measurement Error in Race/Ethnicity

A concern that has arisen with the estimation of mortality change in the United States among white and black groups is that reporting patterns for Hispanic identity may have changed over recent decades. The Hispanic population of the U.S. has higher in- and out-migration, making mortality estimates for this group more difficult to measure (Markides and Eschbach, 2005). Hispanics also have considerably lower mortality rates than other groups (Palloni and Arias, 2004; Markides and Eschbach, 2005). If patterns of Hispanic reporting change over time, and especially if they change differentially across the Current Population Survey and the Vital Statistics databases, then estimates of mortality change for non-Hispanics could be biased.

In this section, we consider and rule out two alternative hypotheses for the measured rise in non-Hispanic white mortality. First, survey questions change subtly over time; for example, the Census permitted people to check multiple race boxes starting in 2000 (Currie, 2018). We show that there are no discontinuities in the combined population records for non-Hispanic white or black populations in our sample, suggesting multiple race reporting is not substantially biasing mortality estimates.

Second, populations' attitudes about their own racial/ethnic identity may change over time. The same person might be more likely to report herself as a given race/ethnicity in 2018 than in 1992. Because Hispanics have lower mortality rates, if white Hispanics are more likely to report Hispanic identity over time in surveys, this phenomenon could bias the non-Hispanic mortality trend upward. We conduct a bounding exercise that shows that even if an implausibly large number of whites changed their identity to Hispanic, there would still be large mortality gains among less educated non-Hispanic whites.

Finally, average misalignment between Hispanic identification on death records and in census counts would not bias mortality estimates unless the error rate changed between 1992 and 2018. Further, researchers have examined potential misreporting of Hispanic identity on death certificates, and have found that reporting of identity on death certificates is in fact reliable and consistent with census data, with error rates consistently falling below 10%, which is too low to explain the patterns described in this paper (Rosenberg et al., 1999; Arias et al., 2008, 2010; Ruiz et al., 2013).

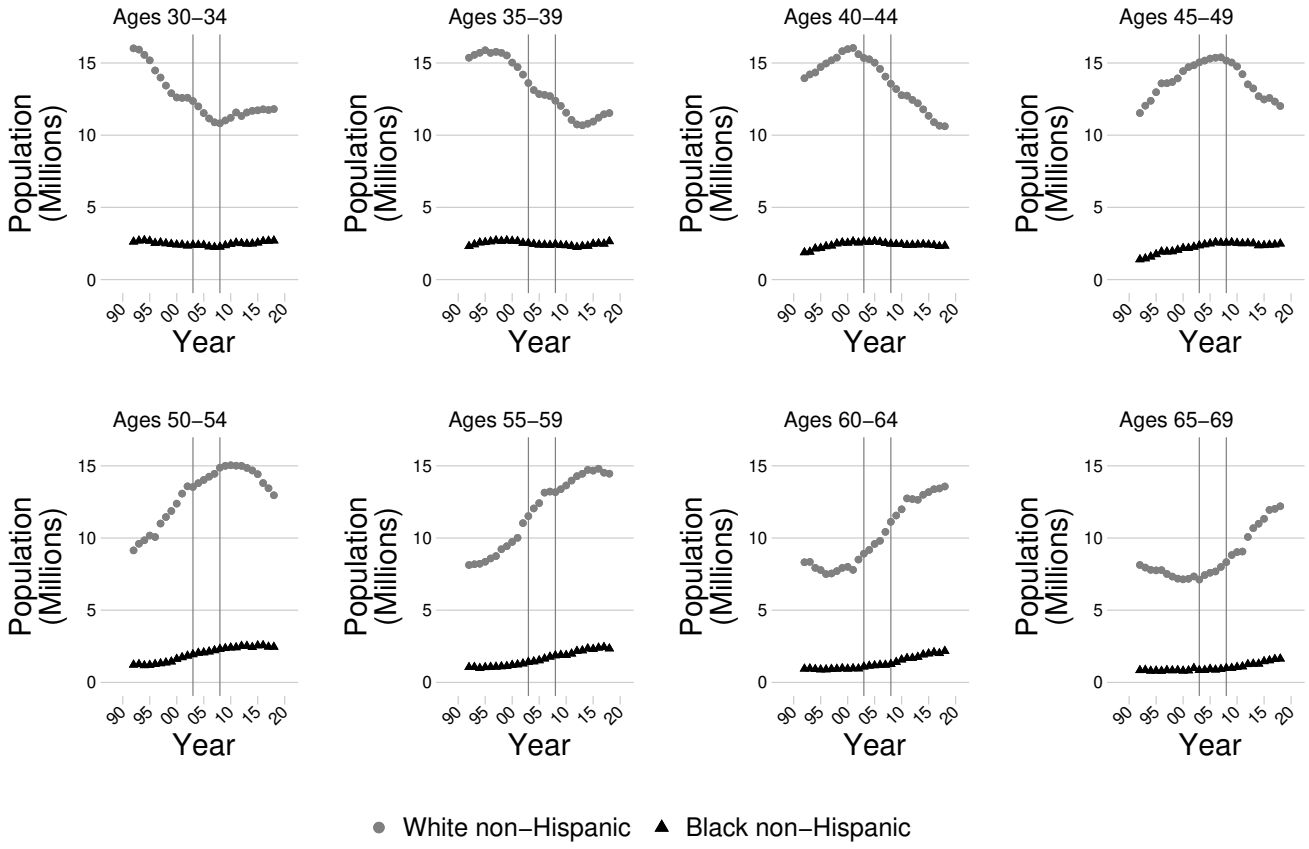
Changes in survey questions. The option to check more than one race box in the 2000 Census has the potential to change the share of the population that reports as any given race (Currie, 2018).

The CPS question on race changed in 2003, while the question in the ACS (which we use only to calculate the institutionalized population) changed in 2008. Figure D4 plots the total population in our dataset by age and race. There is clearly no large discontinuous change in the population of any group in either of these years, suggesting that the option to check multiple races in these national surveys cannot explain the secular trend in rising mortality among the least educated white non-Hispanics. Note also that only 5% of white 20-year-olds in the 2010 Census report multiple races (Currie, 2018). This number drops to less than 2% among white 50-year-olds. This is thus unlikely to be a major concern.

Differences between the Census/CPS and NCHS reporting. A separate concern is that Hispanic identity could be reported differently on death certificates and in Census counts. Misreporting of ethnic identity on death certificates on average would not affect our findings on mortality *changes* unless the frequency of misreporting changes substantially during the sample period. For instance, to create an upward bias in mortality change among non-Hispanic whites, Hispanic identity could be reported correctly on death certificates in 1992 but substantially underreported in 2018. In that case, there must be only small changes in accuracy of reporting in the CPS.

To test the extent to which changes in reporting of Hispanic identity could influence mortality estimates among non-Hispanic white mortality, we simulated misreporting in the data, focusing on non-Hispanic white women aged 50–54 in the least educated 10%. Specifically, we assumed that X% of white Hispanics in 2016–18 would have reported themselves as non-Hispanic white in 1992–94. We therefore reassigned X% of white Hispanic deaths to be counted as white non-Hispanic deaths in 2016–18, and then recalculated bounds on mortality change from 1992–94 to 2016–18. Panels A and B of Figure D5 plot the results of this exercise for non-Hispanic white women and men respectively. Even in the extreme case where 20% of white non-Hispanic female deaths in 2016–18 would have been reported as Hispanic deaths in 1992–94, we would still detect an increase in mortality of [397, 590] deaths per 100,000 among the bottom 10%, putting the lower bound on mortality increase at 65%. This example shows that even an extreme amount of misreporting could not drive our findings for the least educated group; however, note that estimates suggest the true amount of this kind of misreporting is much smaller than this extreme case (Rosenberg et al., 1999; Arias et al., 2008, 2010; Ruiz et al., 2013).

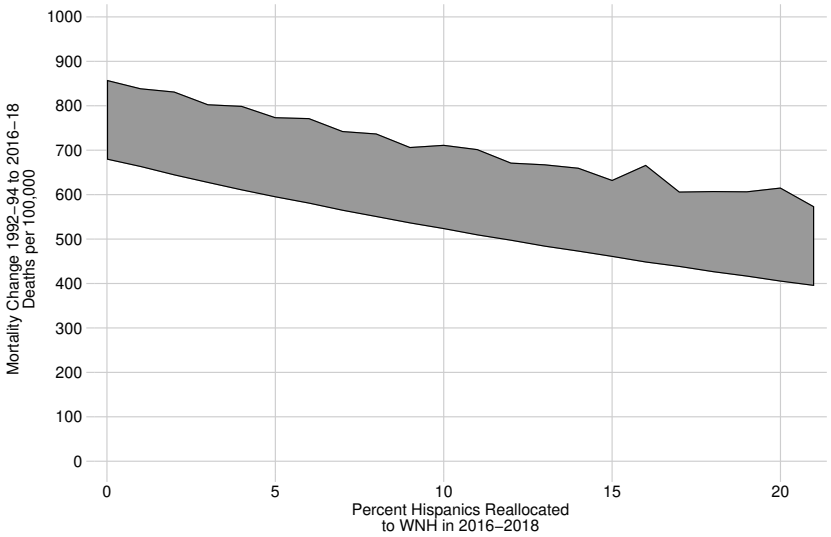
Figure D4
Population Counts by Group



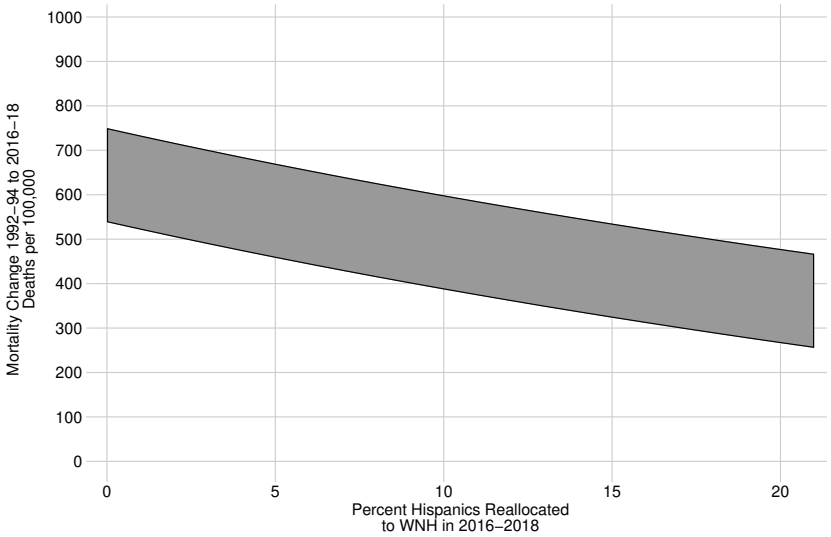
Note: the figure shows the population in each 5-year age bin of black and white non-Hispanics according to the CPS / ACS over the study sample period.

Figure D5
Mortality Changes with Simulated Measurement Error

Panel A: Non-Hispanic White Women Ages 50–54



Panel B: Non-Hispanic White Men Ages 50–54



Note: The figure displays the sensitivity of mortality estimates to measurement error in ethnicity. The figure shows how primary mortality change estimates change if we recode white Hispanic deaths in 2016–18 to white non-Hispanic deaths, leaving reporting in 1992–94 and population totals unchanged. The X axis shows the percentage of Hispanic deaths recoded as non-Hispanic deaths under each scenario. The Y axis shows bounds on mortality change from 1992–1994 to 2016–2018 among non-Hispanic white men and women aged 50–54 in the least educated 10%, under each different recoding. Bounds are otherwise calculated as in the body of the paper.

D.5 Analysis of Measurement Error Using Synthetic CPS Cohorts

This section conducts an exercise to determine whether misreporting of education can explain our estimates of rising mortality among the least educated whites. Specifically, we examine whether changes in CPS cohort sizes are consistent with our mortality estimates. For instance, if we conclude from vital statistics that 5% of dropouts in a certain cohort die in a given year, then we expect that cohort to shrink by about 5% in the following year, after adjusting for other factors that can change cohort size. This approach addresses the concern that if CPS respondents increasingly inflate their level of education (*i.e.* report that they have completed high school when in fact they did not), then the denominator of the mortality rate (the estimated number of high school dropouts) would be increasingly biased downward over time, causing us to overestimate mortality change.

We will put an upper bound on this source of bias by studying the size of a synthetic cohort of dropouts and high school completers in the CPS. The size of a cohort of white high school dropouts can change over time for five reasons: (i) deaths; (ii) migration; (iii) continuing education; (iv) false reports of continuing education; or (v) some report that they are Hispanic, even though they would not have reported this in the past. We calculate an upper limit on the share of individuals who are exiting the sample for reporting reasons; this gives us a bound on the combined bias caused by misreporting of education and ethnicity status in the mortality rate.

We measure the death rate in every period, and we assume that net migration of non-Hispanic middle-aged whites is small enough to ignore. We also estimate the number of individuals passing the GED, which is the primary form of continuing education for individuals who did not complete high school. We obtained the number of GED passers from 1992 to 2013 from the GED Testing Service (American Council on Education, 1993–2014).[42] The number of passers is disaggregated by 5- or 10-year age group; the GED Testing Service also reports the share of passers who are female and the share who are white. The number of passers is not further disaggregated either into single-year age bins or bins describing age * female or age * race. We therefore assume that the number of passers is distributed uniformly across ages within age bins, and that the female and white shares are the same at all ages. We expect that these assumptions will bias downward our estimates of white female passers at

[42]For a sample report, see the 2009 GED Testing Program Statistical Report, which we downloaded from https://files.eric.ed.gov/fulltext/ED512301.pdf. Due to the absence of more recent data, we use 2015 as a terminal year in the CPS and assume the number of GED passers in 2014 and 2015 is the same as in 2013. Alternate assumptions do not materially affect the results.

higher ages, as women may be more likely than men to delay continuing education due to pregnancy.

Given our estimate of the number of GED passers and our estimate of the mortality rate, we can predict how the size of a synthetic CPS cohort of high school dropouts will evolve over time. Any discrepancy between our predicted cohort size and the actual cohort size will be driven by migration (which we expect is small), continuing education in a form other than the GED, false reporting of continuing education, or change in reporting of ethnicity. The discrepancy is therefore an upper bound on the mismeasurement of mortality due to individuals exiting the sample due to misleading reporting of education or ethnicity.

Figure D6 presents the results. The top left panel shows the analysis for non-Hispanic white female dropouts. The gray squares show the CPS population of white female dropouts in the 1950–54 birth cohort, approximately the middle 5-year birth cohort in our study. Their population falls over time due to the five factors above. The black dot-dash line is a linear trend fit to the gray points to eliminate year-on-year noise. The dashed blue line shows how the size of this cohort would have evolved over time from mortality alone, beginning at the CPS trend line in 1992, based on our estimates from the NCHS. The red line shows how this cohort would have evolved when we count deaths and GED completions. In 2015, the gap between the linearized CPS cohort size and the predicted cohort size is 8.0%. Assuming that migration in this cohort is small, this gap is an upper bound on the error in our population count that arises from false reporting of high school completion or changing in reporting of Hispanic ethnicity.

The remaining panels of the figure show the same result for white men, and for white women and men in the 1960–64 birth cohorts. Continuing education explains more of the change in cohort size for the younger 1960–64 birth cohort because individuals are more likely to complete GEDs at younger ages. The potential biases for men are smaller than for women. The potential bias is highest for women in the 1960–64 birth cohort, with a discrepancy of 23.6% between the CPS population and our predicted population. One factor that could explain this discrepancy is the possibility that women are more likely to take the GED later in life than men because of early pregnancies. Our GED passing numbers did not report age profiles for men and women separately, so we had to assume that men and women are equally likely to take the exam in their thirties and forties. If men are more likely to take the exam in their teens and twenties and women are more likely to take it later in life, then our predicted measures would be even closer to the true series for both men and for women.

Nevertheless, for the sake of argument, we can consider how our mortality change measures would

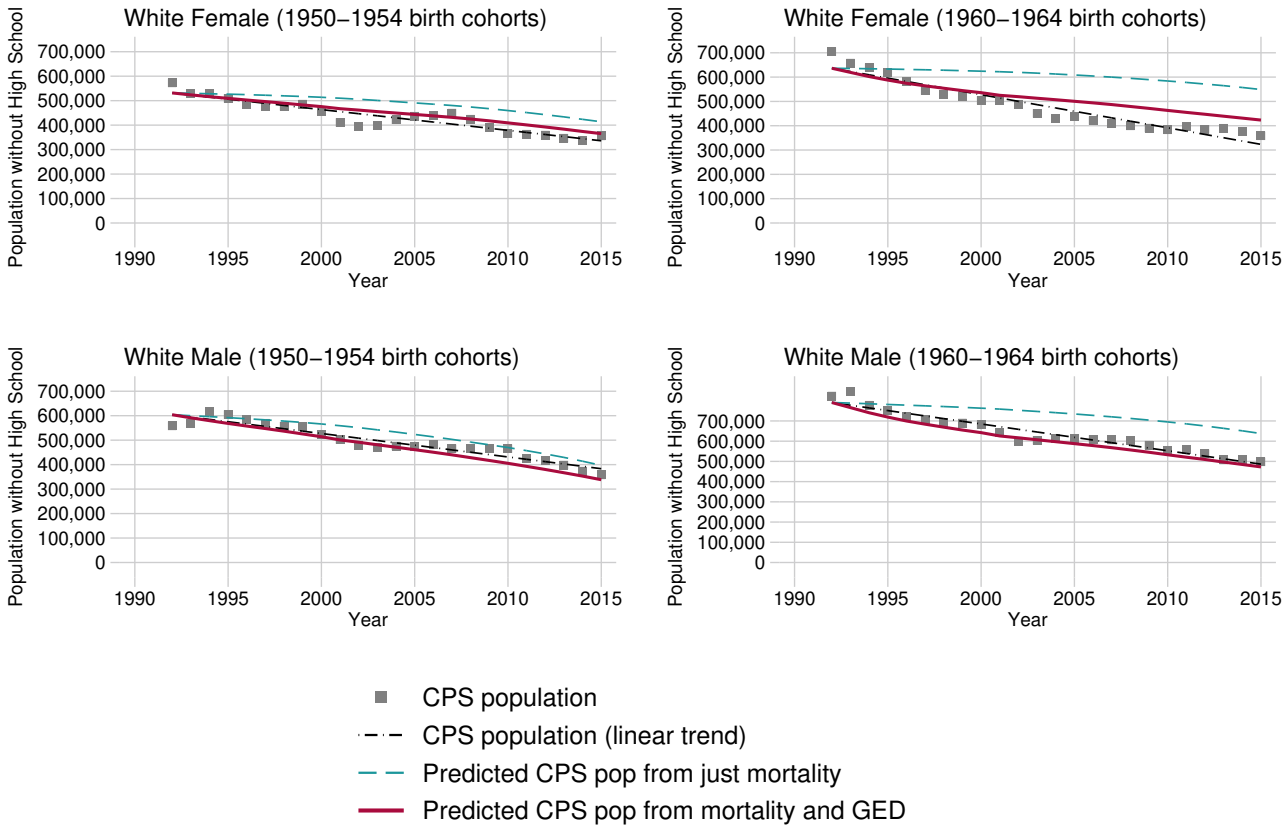
change if our mortality estimates in 2016–18 are biased upward by the worst case estimate of 23.6%.[43] We estimate that 50–54-year-old women (corresponding to the 1960–64 birth cohort in 2016–18) in the least educated 10% experienced mortality increases of 100–150% from 1992–94 to 2016–18. If we underestimated the population of white female dropouts in this birth cohort by 23.6% in 2016–18, the corrected mortality change would be an increase of 62–102%. While this number is smaller, it is still considerably higher than the mortality change estimate in the next education percentile group— among similarly-aged women in the 10th to 45th percentiles, we estimate mortality change between -6% and +20%.

These worst case assumptions are implausible for three reasons. First, the calculation above uses the discrepancy for the 1960–64 birth cohort, which is among the largest in our sample; among women in the 1950–54 birth cohort, assuming the maximum bias would bring mortality change down only from 132–149% to 115–131%. Among men, the bias is less than 10% across all age groups. Second, as noted above, women may be disproportionately likely to complete the GED at higher ages, bringing down the potential bias. Last, there are other mechanisms for individuals to obtain continuing education after dropping out of high school; we are only able to count the GED, and thus incorrectly attributed other schooling to bias. All these reasons suggest we have overstated the potential bias. Nevertheless, even if we make these worst case assumptions, the overall finding of disproportionately rising mortality among the least educated holds up.

Finally, Figure D7 shows similar graphs documenting the change in the size of synthetic CPS cohorts of white high school completers. For these groups, we can only predict population change due to mortality, because we could not obtain population counts of the number of whites obtaining any sort of 2- or 4-year degree by year and age. Nevertheless, even without counting continuing education, the potential discrepancies are very small and within the range that could be plausibly explained entirely by continuing education. In short, false reporting of education in the CPS cannot come close to driving the main results of substantial mortality increases among the least educated whites.

[43] We use 2016–18 as the comparison period, even though we only calculated bias up to 2015 due to availability of the GED data.

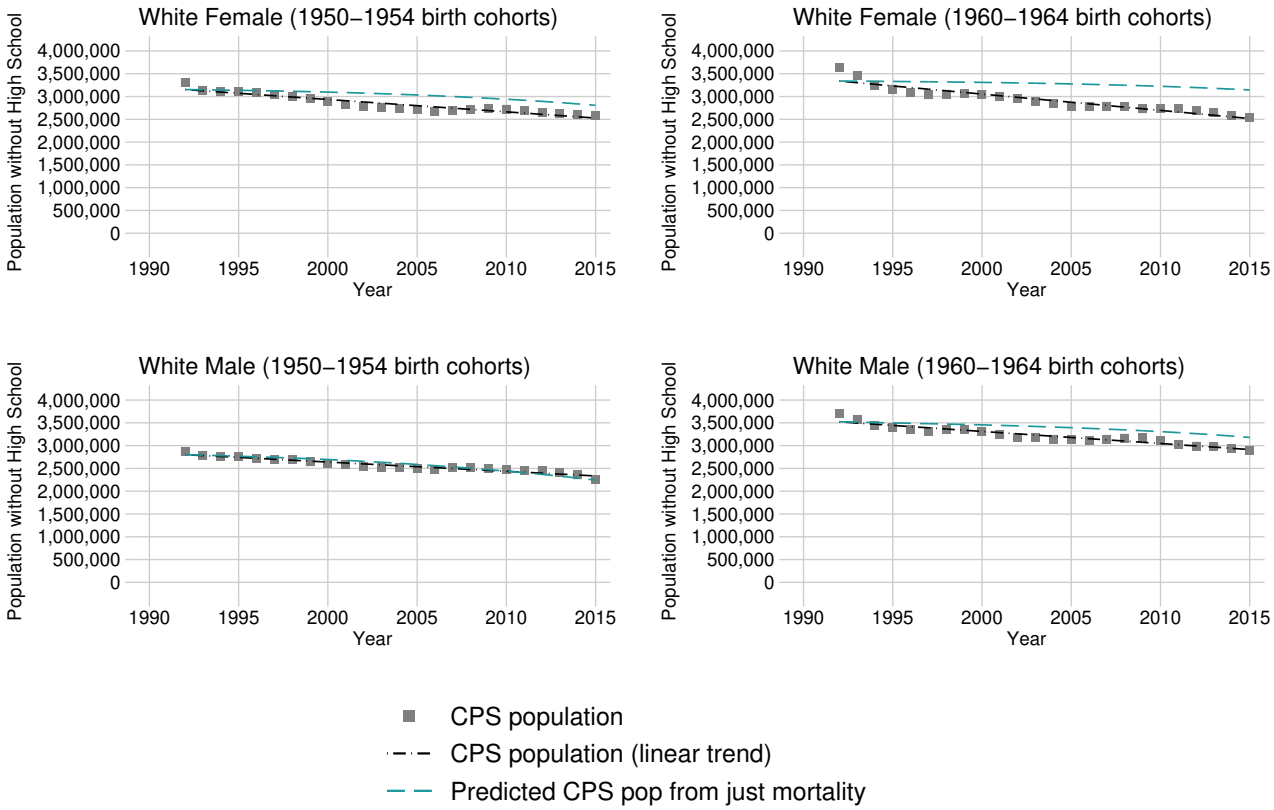
Figure D6
Bounding Measurement Error in CPS Dropout Counts



The figure displays the population in selected cohorts of CPS dropouts. These are compared with the predicted population based on measured mortality and GED completion. The gray points show the counts of members of each gender/education group in each round of the CPS. The dot-dash black line is a linear trend fit to the gray points to eliminate year-on-year noise. The dashed blue lines begin at the CPS trend line in 1992, and show how the CPS population would have evolved from mortality alone. The solid red line shows how the CPS population would have evolved from mortality and GED completion only.

Figure D7

Bounding Measurement Error in CPS High School Completer Counts



The figure displays the population in selected cohorts of CPS high school completers. These are compared with the predicted population based on measured mortality. The gray points show the counts of members of each gender/education group in each round of the CPS. The dot-dash black line is a linear trend fit to the gray points to eliminate year-on-year noise. The dashed blue lines begin at the CPS trend line in 1992, and show how the CPS population would have evolved from mortality alone.

D.6 Comparison with National Health Interview Survey

In this section, we use an alternate data source to validate the findings in the body of the paper. We use the National Health Interview Survey (NHIS), which is an annual repeated cross-section survey of about 35,000 households and 87,000 individuals. Mortality rates can be calculated directly from the NHIS, because NHIS records are intermittently linked to death certificate records from the National Death Index. We can therefore estimate a population subgroup mortality rate as the share of individuals who are deceased in a given followup period.

Because mortality rates and education/ethnicity are all measured for the same individuals, there is no possibility of inconsistent measurement of ethnicity or education between death counts and population counts. For instance, if some individuals without a high school education report that they have a high school education, both their deaths and their population will be counted among the high school group. This may create a small bias if mortality is correlated with misreporting, but it will be considerably less bias than if their deaths are counted in the dropout group and their population is counted in the high school completion group. This said, NHIS-based measures of mortality slightly underestimate aggregate mortality relative to vital statistics data, especially for older white women, because the NHIS sample population appears to be more healthy than average (Ingram et al., 2008).

We obtained public NHIS data with mortality followup for NHIS participants linked to death records through 2015. NHIS interviews occur throughout the year.[44] We aggregated results across all ages using the standardized U.S. population distribution, as in Section 4.4. Because our aim in this section is to validate the raw mortality estimates from the NCHS, we present raw mortality change for education levels, rather than bounding the mortality change in constant education percentiles.

Figure D8 below compares estimates of annualized mortality change in the NHIS vs. our estimates from the NCHS for the four key groups in our study: non-Hispanic white female dropouts and high school completers, and non-Hispanic white male dropouts and high school completers. Mortality change in the NHIS is calculated from the average mortality in periods 1997–1999 to the last period in which the n-year mortality rate can be calculated. For example, we can compute 1-year mortality for the 2014 data, 2-year mortality for the 2013 data, etc.

The red points in Figure D8 show NHIS mortality estimates with 95% confidence intervals. Even

[44]The sample of deaths in the year following the NHIS survey is extremely small. In some of our subgroups, there were zero deaths reported in the shortest followup periods.

with the 6-year followup period, the NHIS sample is too small to precisely estimate mortality change over the sample period. Most point estimates are very close to our mortality change estimates from NCHS data (indicated by the dashed green line), but in many cases the NHIS confidence interval includes both zero and our measure. In almost all cases, our NCHS measure of mortality change is within one standard error of the NHIS estimate and they are particularly close for the largest sample mortality followup in NHIS (6 years).[45]

Health status. NHIS estimates of mortality are imprecise because the number of middle-aged white dropouts in the sample who die is very small. We can obtain more precise estimates of health status, which is reported by all respondents. NHIS respondents are asked to report their perception of their own health on a five point scale, where one reflects very good health, and five reflects very bad health. In Figure D9, we show the annualized change in self-reported health status from 1997 to 2014 at each age and education level, plotted with a lowess smoother.

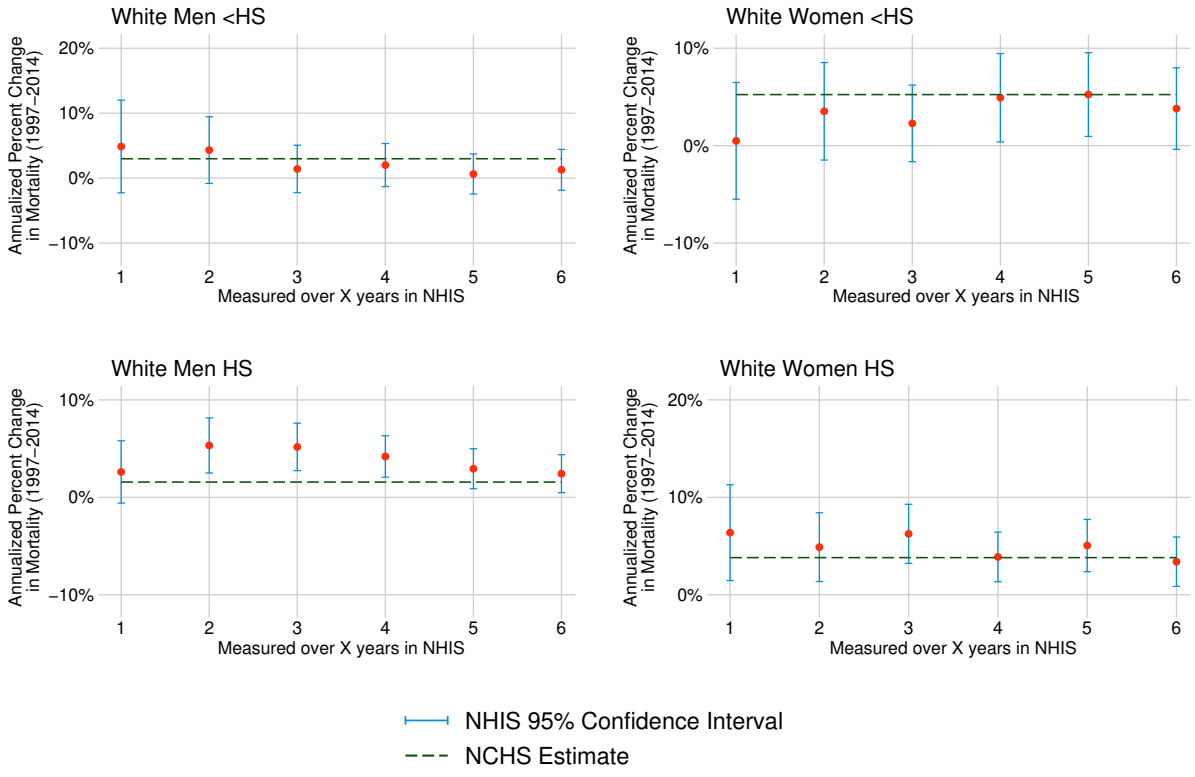
The left panel shows the result for non-Hispanic white women. At most ages, dropouts have experienced worse health declines than members of any other group. Among 40–60-year-old women, self-reported health status has gotten 0.0175 points worse *per year*, or 0.2 points worse on a 5 point scale from 1997 to 2009. The age pattern of the health decline closely matches our mortality results in Figure 6, with the greatest divergence between dropouts and high school graduates occurring between ages 40 and 60. We observe less change in the education-health gradient among men, but dropouts suffer the worst relative deterioration in health between ages 50–70, corresponding exactly to the ages where we document the greatest differential increases in mortality among the bottom 10% in Figure 6.

In conclusion, measures of mortality and self-reported health in the NHIS are broadly consistent with our findings of increased mortality among middle-aged whites in the least educated 10%. These NHIS measures are less precise than the vital statistics mortality data, but they do not suffer from any division bias that could be caused by changes in how individuals respond to questions about their education or ethnicity over time.

[45]Compared with our results, NHIS calculates slightly higher mortality increases for high school completers and slightly lower for dropouts, but the opposite interpretation also falls within the confidence interval. The NHIS estimates are just not precise enough to make any clear statement about the difference between these groups.

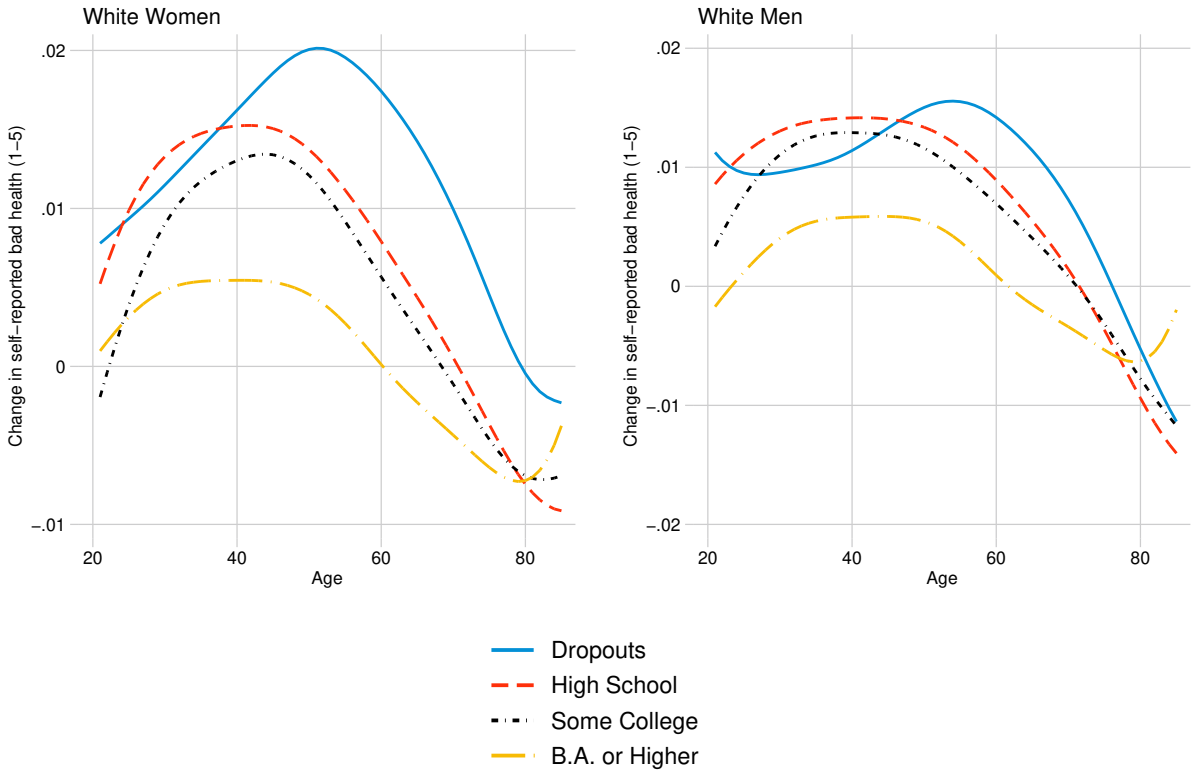
Figure D8

Annualized Mortality Change Estimates from NHIS and from NCHS (1997–2015)



The figure compares estimates of annualized mortality change in the National Health Interview Survey (NHIS) with vital statistics data from the NCHS, used in the body of the paper. The annualized change in the mortality rate from 1997–99 to 2015 according to NHCS (vital statistics data) is indicated by the dashed green line. The six different points for each panel reflect annualized mortality changes in the NHIS based on measurement of 1-year mortality, 2-year mortality, 3-year mortality, and so on. 1-year mortality is reported in the NHIS using 1997–1999 as a base period and comparing to participants interviewed in 2014 with mortality followup to 2015. 2-year mortality change is reported from 1997–1999 and comparing to 2013 participants, and so on. These different estimates are nevertheless comparable because the mortality change estimates are annualized. Standard errors are calculated for the NHIS using NHIS sample weights.

Figure D9
Change in Self-Reported Health Status (NHIS, 1997–2015)



The figure shows a lowess fit to the annualized change in self-reported health from 1997 to 2015 at each age, according the NHIS. Self-reported health status is on a five point scale, where one represents the best health and five represents the worst. We calculate annualized change for each age cohort / education group by regressing self-reported health status on a year variable. We then plot the year coefficients for each age/education group using a lowess smoother to minimize noise across single-year age cohorts. The series therefore shows the predicted change in self-reported health for an individual in a given age, gender and education group.

D.7 Measurement Error in Education

This subsection further explores the possibility that measurement error in education biases the main finding of rising mortality among the least educated whites. High school graduation rates on death certificates are thought to be inflated on average, though higher education levels appear to be reported accurately (Sorlie and Johnson, 1996). There is no evidence that this bias has changed during the sample period, but (as discussed in Appendix D.5) if overreporting of high school completion on death certificates (but not in the CPS) were to decline over time, it would cause mortality change among high school dropouts to biased upward.

For measurement error to explain the rising mortality among dropouts described in this paper, it would need to be the case: (i) that the measurement error in education changes substantially during the sample period; (ii) that measurement error changes differentially for deaths of deaths of despair and deaths in other categories, at different ages, but changes considerably less for deaths of despair; and (iii) that measurement error has declined dramatically for non-Hispanic whites, but has changed only minimally for blacks. Misreporting rates would have to differ substantially across age groups as well—for example, we find that dropouts account for nearly all mortality gain among 50–54-year-olds white women, but that rising mortality is equally distributed among dropouts and high school graduates among 25–29-year-old white women.

In contrast, if white respondents were increasingly inflating their education in the CPS, we would expect a uniform increase in mortality from all causes, across all ages, which is not remotely what we find. Note also that other researchers have noted rising health disparities between dropouts and high school graduates, in datasets where misreporting of education is unlikely to be a concern (Montez et al., 2011).

We show here that even if one treats the distinction between dropouts and high school graduates as unreliable, the central finding of rising mortality among the least educated remains clear. To show this, we combine dropouts and high school graduates into a single education group, and we estimate mortality change for three constant education percentile bins: (i) percentiles 0-45 (corresponding approximately to dropouts *and* high school graduates in in the middle of the sample period; (ii) percentiles 45-70; and (iii) percentiles 70-100. Figure D10 shows estimates of mortality change from 1992–94 to 2016–18 for these three groups. We continue to find a dramatic rise in mortality among whites at the bottom of education distribution, though we have defined the bottom more broadly

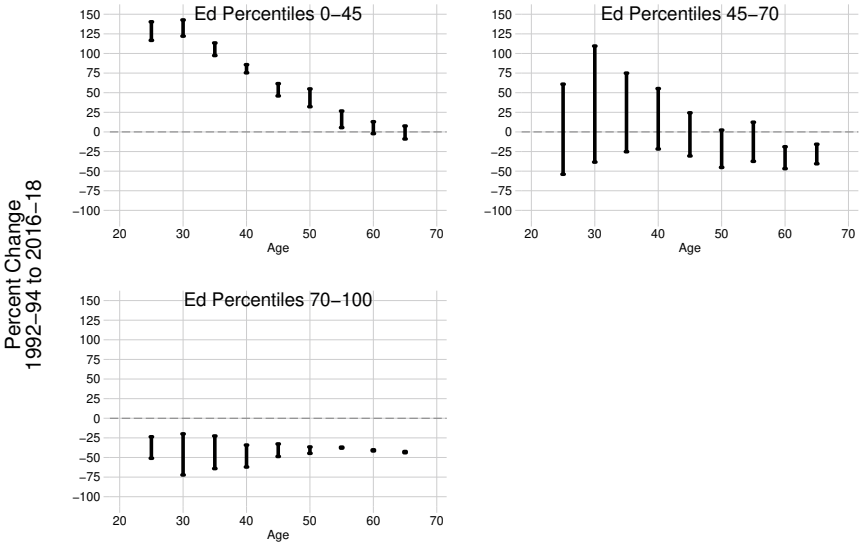
here. We find that mortality among the bottom 45% of the education distribution rises over 50% for white women below 45, rises 0–50% for women 40-59, and is flat at older ages. Mortality among white men in the bottom 45% is rising for age groups below 60.

In short, we find significant increases in mortality disparities by education even when we pool high school graduates and high school dropouts. However, by ignoring the distinction between dropouts and high school graduates, we miss the important difference in causes of death noted in Figure 7—mortality increases outside of the very bottom of the distribution are driven primarily by deaths of despair, but mortality increases in the bottom 10% are driven by a multitude of causes. Absent evidence that measurement error in death records has changed substantially during the sample period, rising mortality among the least educated 10% of the population should be taken seriously.

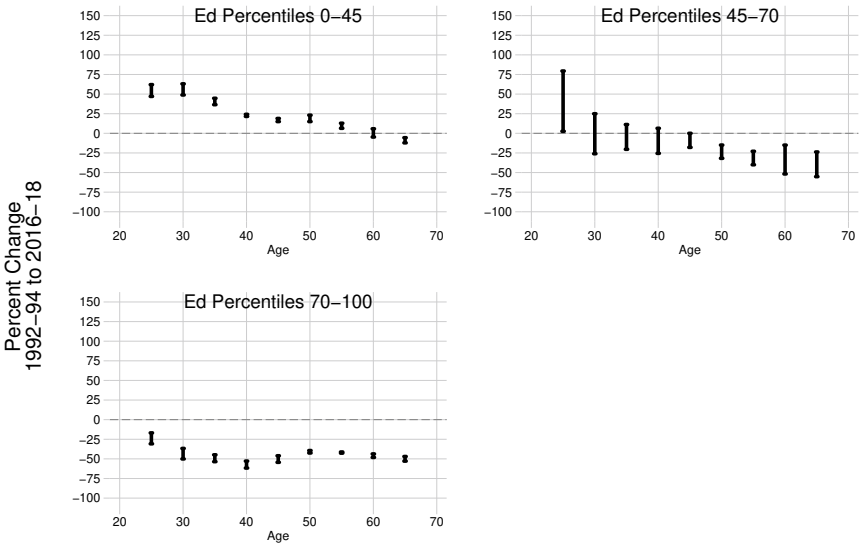
Figure D10

Change in non-Hispanic White Mortality: 3 Education Groups

Panel A: Non-Hispanic White Women



Panel B: Non-Hispanic White Men



Note: The graph shows changes in mortality by age, sex, race, and constant percentile education bin. The figure is analogous to Figure 6, but with the bottom two education categories (percentiles 0–10 and 10–45) pooled into a single category covering percentiles 0–45. The lines show the bounded set containing the percentage change in the mortality rate from 1992–1994 to 2016–2018 for the given group. Bounds are calculated following Section 3. Sources: CPS, NCHS.