Methodological Appendix
Computation of Life Expectancy Values for Exact Quantiles of Education

This appendix describes in detail how to implement our demographic method to compute mortality or life expectancy values for any given quantile/percentile of the education distribution. For ease of exposition, we will assume in this appendix that the reader wishes to compute age-specific mortality rates corresponding to each quintile (20% groupings) of the education distribution. If an analyst instead prefers to compute mortality rates corresponding to some other education quantile, they need only vary their approach in Step 5 listed below.

Step 0: *Acquire data on education-specific mortality and the education distribution for each age group*

This step is required prior to implementing the method. The basic input data for the method includes age-specific mortality rates for each education category and the distribution of population across the various education categories within each age group.

The running example in this appendix will use simulated data on a population consisting of nine education categories that vary in size across age groups. The nine education categories are people who completed at most: primary school, intermediate school, junior high school, some high school without a diploma, high school with diploma, some college without a degree, a bachelor's degree, a master's degree, or a Ph.D. or other doctoral degree. We use data on five-year age groups ranging from 30-34 years to 80-84 years and an open-ended age group for people aged 85 or older. The education distribution for each age group is shown in Figure 1 and arrayed in Table 1. Each stacked bar corresponds to an age group and sums to 1. Figure 1 shows that this population likely experienced an educational expansion wherein the lower education categories shrunk in relative size across birth cohorts while some of the higher education groups grew in relative size for more recent cohorts.

In practice, these age-specific education distributions can be computed using either current data on educational attainment or using data on the educational attainment for each birth cohort at some earlier age (e.g., the education distribution of people aged 50-54 in 2015 could come from data on the educational attainment of people aged 30-34 in 1995). The benefit of using current data is that it is easier to acquire and incorporates any migration into or out of the population over the past decades. The benefit of using data on cohorts at an earlier age is that it is less subject to education misreporting and potentially less affected by mortality selection. Except in rare cases of very severe mortality selection or very high migration rates, both education distribution measures should yield qualitatively similar results.

The other type of input data is age-specific mortality rates for each education category. In our example, we have nine education categories and 12 age groups, so we have a corresponding 9x12 matrix of death rates, as shown in Table 2. The period life expectancies at age 30

corresponding to these education categories are shown at the bottom of Table 2. The mortality rates are plotted on linear and logarithmic scales in Figure 2. That figure shows the very stark differences in mortality rates across the nine education categories. The life expectancies at age 30 range from 37.8 years for the bottom education category to 61.7 for the top education category. The life expectancy at age 30 for the overall population lies in between at a value of 46.3 years.

Step 1: *Compute the ratio of age-education-specific mortality rates to the overall (non-education-specific) age-specific mortality rate*

In the main part of the paper we establish a convex (negative logarithmic) relationship between relative education and relative mortality. Step 1 is thus to compute relative mortality. This step is fairly straightforward: for each age-education group, we compute the ratio of mortality to the age-specific mortality rate for the overall population. For our example, we compute the ratio of each of the first nine columns in Table 2 to the 'Overall' column in Table 2. The results of this computation are shown in Table 3. The mortality ratios range from a low of 0.368 to a high of 2.364.

Step 2: *Compute the midpoints of the education groups in each age group*

In a given age group, each education category exists along some segment of the education distribution. For example, according to Table 1, for people aged 30-34 the third education category lies between the 13th percentile (0.04+0.09=0.13) and the 23rd percentile (0.04+0.09+0.10=0.23). The midpoint of the third education category thus lies at the 18th percentile ((0.13 + 0.23)/2 = 0.18). We compute these midpoints for each education category in each age group for potential use in Steps 3 and 4. These midpoints are shown in Table 4.

Step 3: *Examine the relationships between the mortality ratios and the education percentiles in each age group to ensure a convex relationship exists*

Because this method relies on the finding that there is a negative logarithmic relationship between relative education and relative mortality, we must ensure that this relationship holds before proceeding. There are a number of ways one can examine this relationship. Two of the simplest ways to do so are to inspect a graph of the mortality ratios (from Step 1) plotted against the education midpoints (from Step 2) or to regress the natural logarithm of the mortality ratios against the education midpoints for each age group and inspect the R-squared values.

Using the data from our running example, visual inspection shows an approximate logarithmic relationship for all age groups. Regressions of the natural log of the mortality ratios on the education midpoints show R-squared values of 94% or higher for each age group. The combined data for all age groups is plotted in Figure 3. Our inspection of age-specific graphs and the R-squared values suggest that our assumption of a convex relationship holds, and we can proceed with our analysis. Note that in practice, one should not always expect to find such high R-squared values and the researcher will need to make an individual judgment about whether this convexity assumption is met.

Step 4: *Estimate the values of the α and β parameters in the equation $ln(R_i) = \alpha + \beta e_i$ for each age group*

This method assumes that there is a negative logarithmic relationship between relative mortality and relative education for each age group. This relationship is summarized by two parameters, α and β, in the equation $\ln(R_i) = \alpha + \beta e_i$ where R_i is the mortality ratio value and e_i is the relative education value. In our example, we have nine education categories, so we can identify the α and β values by regressing the natural logarithm of the nine mortality ratios on the education midpoints for the nine education categories. We run this regression for each of the 12 age groups to get twelve α and twelve β estimates.

The more education categories in the data, the better will the regression approach work. If there are either very few education categories or if one of the categories is very large, the regression approach may not be consistent. In this scenario, we suggest that researchers instead use maximum likelihood estimation or calibration based on equation (3) in the main text or a similar method to compute α and β estimates. In general, though, we suggest using the regression approach for its computational simplicity.

Step 5: *Compute the average mortality ratio values for the quantiles of interest for each age group*

Now that we have all of the sets of α and β values for each age group, we can compute the average mortality ratio corresponding to any given education quantile. In our example, we want to compute mortality rates for education quintiles (20% groupings). We thus use the following equation to compute the average mortality ratio ($\bar{R}_{(a,b)}$) for each of the five quintiles in each age group:

$$\bar{R}_{(a,b)} = \frac{e^\alpha}{\beta(b-a)}\left(e^{\beta b} - e^{\beta a}\right).$$

For the first quintile, a=0.00 and b=0.20. We thus plug in those values for a and b in the above equation and, for each age group, we plug the corresponding α and β estimates into the equation. We repeat the same procedure for the subsequent quintiles in each age group (so for the second quintile, a=0.20 and b=0.40, for the third a=0.40 and b=0.60, and so on). The α and β values remain the same across quintiles and only vary across age groups. This exercise produces an average mortality ratio value for each education quintile in each age group. Average mortality ratios for the age-specific education quintiles are shown for our example in the middle five columns of Table 5.

Since the quantiles of interest in this example are education quintiles, we use the aforementioned a and b values in our computations. For researchers interested in other quantiles, they can plug in different a and b values to compute average mortality ratios of interest. For example, if a researcher is interested in computing mortality rates for the bottom 34% of the population and the top 21% of the population, their (a,b) values would be (0,0.34) and (0.79,1.00) for those two groups, respectively.

Step 6: *Multiply the average mortality ratio values for each quantile by the corresponding overall age-specific death rates to get the quantile-specific death rates for each age group*

Now that we have the average mortality ratio values for each quintile in each age group, we simply multiply these values by their corresponding overall age-specific mortality rates. This computation is shown in the final five columns of Table 5. Each of the middle five columns (the mortality ratios) is multiplied by the 'Overall $_nM_x$' (age-specific mortality rate) column to produce education quintile-specific death rates for each age group.

Step 7: *Compute life expectancies corresponding to each quantile*

The final step is to compute life expectancy estimates for our newly-produced education quintile-specific death rates. To do so, we simply apply standard life table techniques to the quintile-specific death rates. In our example, we use graduation to estimate life table $_na_x$ values (Preston, Heuveline, and Guillot 2001). We compute abridged life tables for each of the last five columns in Table 5. The corresponding life expectancies at age 30 are shown in the bottom row of Table 5. They range from 38.9 years to 59.7 years.

Summary

The method outlined above can be used to generate estimates of education-specific life expectancy that are consistent across time and across populations, since the estimates themselves are specific to fixed percentile ranges of the education distribution. Potential applications of this method include comparing educational inequality in mortality or life expectancy across national populations, across racial or ethnic groups, or over different time periods for the same subpopulation. The simplicity of the method combined with its minimal data requirements suggest that it can be easily implemented and should be adopted by researchers seeking to understand the processes underlying changing inequality in mortality.

Figure 1. Education Distribution by Age Group
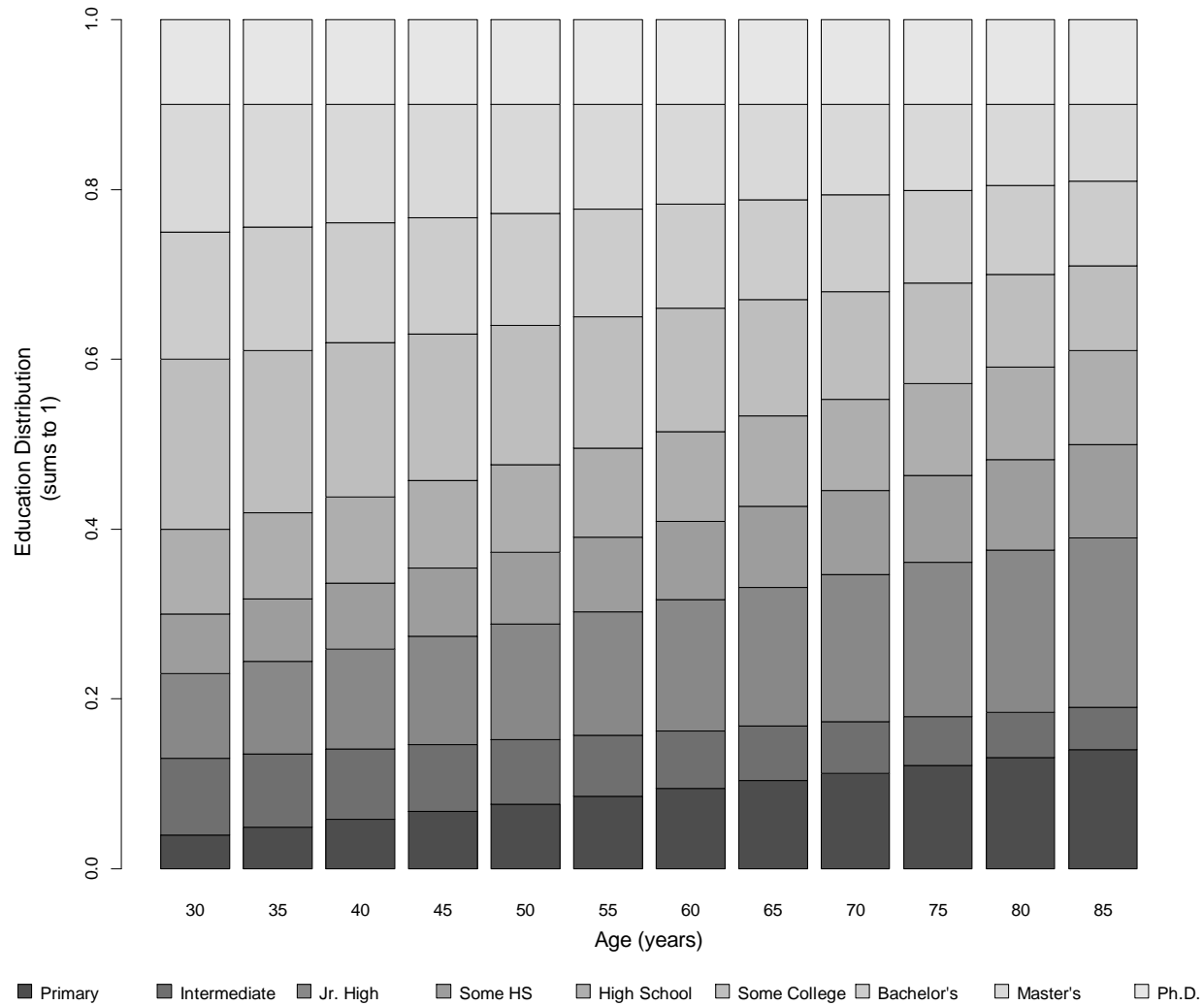
Note: Each stacked bar corresponds to a five-year age group and sums to 100%. Lighter colors represent higher education categories and darker colors represent lower education categories.

Figure 2. Age-Specific Mortality Rates by Education Category

Figure 3. Relationship between Relative Mortality and Relative Education, All Age Groups
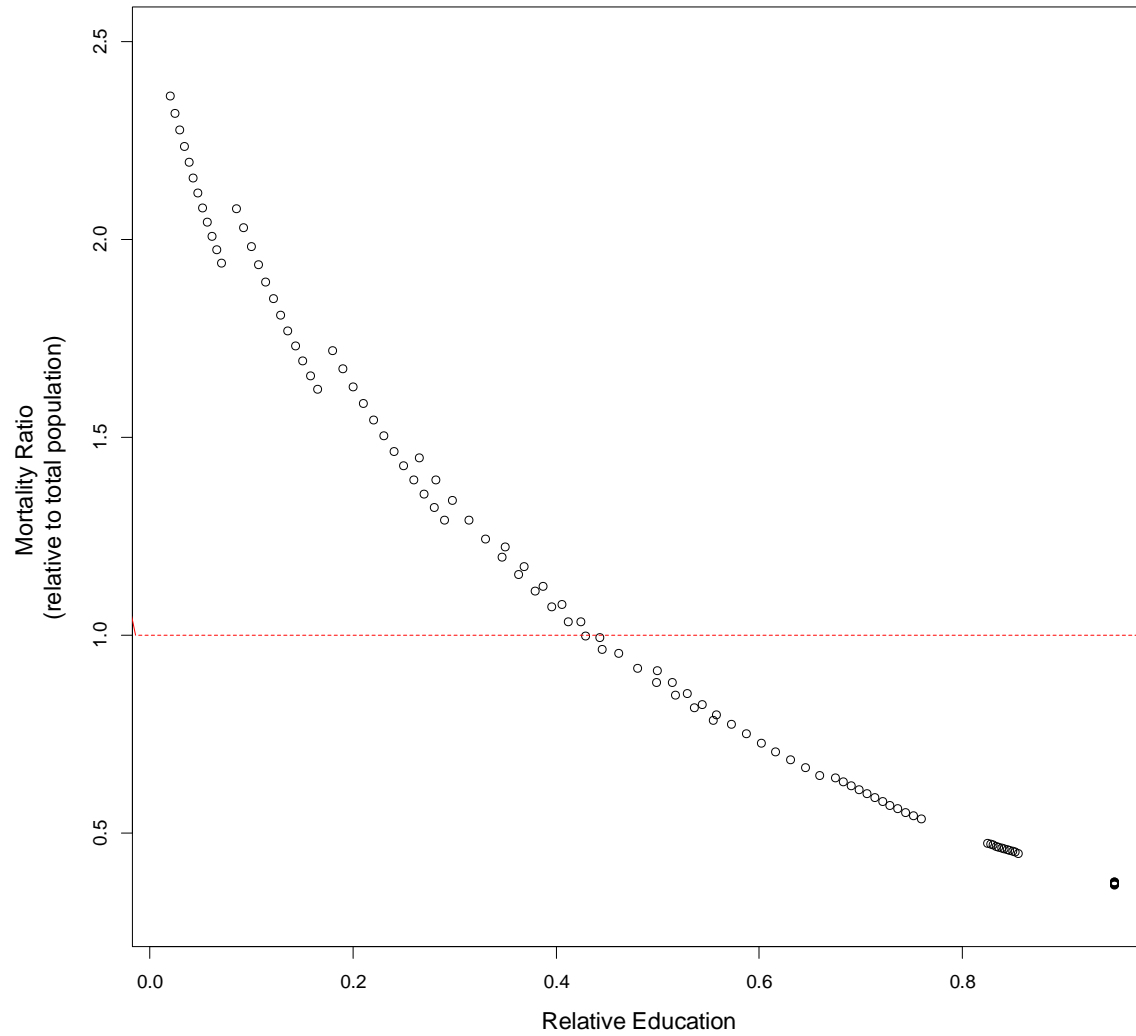
Note: For parsimony, this figure plots the relationship for all age groups at once. However, the convex relationship holds within each age group as well.

Table 1. Education Distributions by Age Group

Age Group	Age-Specific Education Distributions									
	1	2	3	4	5	6	7	8	9	Total
30-34	0.040	0.090	0.100	0.070	0.100	0.200	0.150	0.150	0.100	1.000
35-39	0.049	0.086	0.109	0.074	0.101	0.191	0.145	0.145	0.100	1.000
40-44	0.058	0.083	0.118	0.077	0.102	0.182	0.141	0.139	0.100	1.000
45-49	0.067	0.079	0.127	0.081	0.103	0.173	0.136	0.134	0.100	1.000
50-54	0.076	0.075	0.136	0.085	0.104	0.164	0.132	0.128	0.100	1.000
55-59	0.085	0.072	0.145	0.088	0.105	0.155	0.127	0.123	0.100	1.000
60-64	0.095	0.068	0.155	0.092	0.105	0.145	0.123	0.117	0.100	1.000
65-69	0.104	0.065	0.164	0.095	0.106	0.136	0.118	0.112	0.100	1.000
70-74	0.113	0.061	0.173	0.099	0.107	0.127	0.114	0.106	0.100	1.000
75-79	0.122	0.057	0.182	0.103	0.108	0.118	0.109	0.101	0.100	1.000
80-84	0.131	0.054	0.191	0.106	0.109	0.109	0.105	0.095	0.100	1.000
85+	0.140	0.050	0.200	0.110	0.110	0.100	0.100	0.090	0.100	1.000

Note: The numbers 1 through 9 correspond to each of the nine education categories, with 1 being the least-educated group and 9 being the most-educated group. 'Total' corresponds to the sum of the distribution across education categories within each age group.

Table 2. Age-Specific Mortality Rates by Education Category

Age Group	1	2	3	4	5	6	7	8	9	Overall
					Age-Specific Mortality Rates by Education Category					
30-34	0.00269	0.00236	0.00195	0.00165	0.00139	0.00104	0.00073	0.00054	0.00042	0.00114
35-39	0.00409	0.00358	0.00295	0.00246	0.00207	0.00155	0.00111	0.00083	0.00065	0.00176
40-44	0.00624	0.00543	0.00446	0.00367	0.00308	0.00234	0.00169	0.00128	0.00101	0.00274
45-49	0.00951	0.00824	0.00674	0.00549	0.00459	0.00351	0.00259	0.00198	0.00158	0.00425
50-54	0.01450	0.01251	0.01019	0.00820	0.00683	0.00528	0.00395	0.00307	0.00245	0.00660
55-59	0.02211	0.01898	0.01542	0.01227	0.01018	0.00794	0.00604	0.00474	0.00381	0.01025
60-64	0.03372	0.02881	0.02332	0.01836	0.01519	0.01195	0.00922	0.00732	0.00593	0.01592
65-69	0.05143	0.04375	0.03529	0.02747	0.02266	0.01799	0.01409	0.01131	0.00923	0.02472
70-74	0.07846	0.06644	0.05342	0.04114	0.03383	0.02709	0.02154	0.01747	0.01436	0.03839
75-79	0.11970	0.10091	0.08089	0.06163	0.05052	0.04081	0.03292	0.02700	0.02233	0.05961
80-84	0.18264	0.15331	0.12250	0.09236	0.07548	0.06151	0.05034	0.04172	0.03474	0.09255
85+	0.27872	0.23295	0.18556	0.13848	0.11282	0.09275	0.07699	0.06449	0.05405	0.14370
Life Expectancy at Age 30	37.8 years	39.5 years	42.0 years	45.0 years	47.6 years	50.8 years	54.4 years	58.0 years	61.7 years	46.3 years

Note: The numbers 1 through 9 correspond to each of the nine education categories, with 1 being the least-educated group and 9 being the most-educated group. 'Overall' corresponds to the age-specific mortality rates for the total (not education-specific) population. Life expectancy at age 30 was computed using standard life table techniques and graduation to compute $_na_x$ values.

Table 3. Ratios of Age-Education-Specific Mortality Rates to Overall Mortality Rates

| Age Group | Age-Specific Mortality Ratios | | | | | | | | | |
	1	2	3	4	5	6	7	8	9	Overall
30-34	2.364	2.078	1.719	1.449	1.223	0.911	0.640	0.474	0.368	1.000
35-39	2.320	2.030	1.673	1.393	1.172	0.881	0.629	0.472	0.369	1.000
40-44	2.278	1.983	1.628	1.340	1.124	0.852	0.619	0.469	0.370	1.000
45-49	2.236	1.937	1.585	1.290	1.078	0.825	0.608	0.467	0.371	1.000
50-54	2.196	1.893	1.543	1.242	1.034	0.799	0.598	0.464	0.371	1.000
55-59	2.156	1.851	1.503	1.196	0.993	0.774	0.589	0.462	0.372	1.000
60-64	2.118	1.810	1.465	1.153	0.954	0.750	0.579	0.460	0.373	1.000
65-69	2.080	1.770	1.428	1.111	0.917	0.727	0.570	0.457	0.373	1.000
70-74	2.044	1.731	1.392	1.072	0.881	0.706	0.561	0.455	0.374	1.000
75-79	2.008	1.693	1.357	1.034	0.847	0.685	0.552	0.453	0.375	1.000
80-84	1.973	1.656	1.324	0.998	0.816	0.665	0.544	0.451	0.375	1.000
85+	1.940	1.621	1.291	0.964	0.785	0.645	0.536	0.449	0.376	1.000

Note: The numbers 1 through 9 correspond to each of the nine education categories, with 1 being the least-educated group and 9 being the most-educated group. 'Overall' corresponds to the total (not education-specific) age-specific population, so the mortality ratio for that category is always 1.000.

Table 4. Midpoints of Education Categories in the Education Distributions by Age Group

Age Group	Age-Specific Education Category Midpoints								
	1	2	3	4	5	6	7	8	9
30-34	0.020	0.085	0.180	0.265	0.350	0.500	0.675	0.825	0.950
35-39	0.025	0.092	0.190	0.281	0.369	0.515	0.683	0.828	0.950
40-44	0.029	0.100	0.200	0.298	0.387	0.529	0.690	0.830	0.950
45-49	0.034	0.107	0.210	0.314	0.406	0.544	0.698	0.833	0.950
50-54	0.038	0.114	0.220	0.330	0.425	0.558	0.706	0.836	0.950
55-59	0.043	0.121	0.230	0.347	0.443	0.573	0.714	0.839	0.950
60-64	0.047	0.129	0.240	0.363	0.462	0.587	0.721	0.841	0.950
65-69	0.052	0.136	0.250	0.380	0.480	0.602	0.729	0.844	0.950
70-74	0.056	0.143	0.260	0.396	0.499	0.616	0.737	0.847	0.950
75-79	0.061	0.150	0.270	0.412	0.518	0.631	0.745	0.850	0.950
80-84	0.065	0.158	0.280	0.429	0.536	0.645	0.752	0.852	0.950
85+	0.070	0.165	0.290	0.445	0.555	0.660	0.760	0.855	0.950

Note: The numbers 1 through 9 correspond to each of the nine education categories, with 1 being the least-educated group and 9 being the most-educated group. 'Total' corresponds to the sum of the distribution across education categories within each age group.

Table 5. Estimated Mortality Ratios and Rates by Education Quintile

Age Group	Overall $_nM_x$	$\hat{\alpha}$	$\hat{\beta}$	Education Quintile-Specific Mortality Ratios					Education Quintile-Specific Mortality Rates				
				0%-20%	20%-40%	40%-60%	60%-80%	80%-100%	0%-20%	20%-40%	40%-60%	60%-80%	80%-100%
30-34	0.00114	0.90	-2.00	2.030	1.362	0.913	0.613	0.411	0.00231	0.00155	0.00104	0.00070	0.00047
35-39	0.00176	0.89	-1.98	2.012	1.353	0.910	0.612	0.411	0.00355	0.00239	0.00160	0.00108	0.00073
40-44	0.00274	0.88	-1.97	1.995	1.344	0.906	0.611	0.412	0.00546	0.00368	0.00248	0.00167	0.00113
45-49	0.00425	0.87	-1.96	1.977	1.336	0.902	0.610	0.412	0.00841	0.00568	0.00384	0.00259	0.00175
50-54	0.00660	0.86	-1.95	1.960	1.327	0.899	0.609	0.412	0.01294	0.00877	0.00594	0.00402	0.00272
55-59	0.01025	0.85	-1.94	1.943	1.319	0.895	0.608	0.413	0.01992	0.01353	0.00918	0.00623	0.00423
60-64	0.01592	0.84	-1.92	1.926	1.311	0.892	0.607	0.413	0.03067	0.02087	0.01420	0.00967	0.00658
65-69	0.02472	0.83	-1.91	1.909	1.303	0.889	0.606	0.413	0.04721	0.03220	0.02197	0.01499	0.01022
70-74	0.03839	0.82	-1.90	1.893	1.294	0.885	0.605	0.414	0.07267	0.04969	0.03398	0.02324	0.01589
75-79	0.05961	0.81	-1.89	1.877	1.286	0.882	0.604	0.414	0.11186	0.07668	0.05256	0.03603	0.02469
80-84	0.09255	0.80	-1.88	1.861	1.278	0.878	0.604	0.415	0.17220	0.11832	0.08130	0.05586	0.03838
85+	0.14370	0.79	-1.86	1.845	1.271	0.875	0.603	0.415	0.26509	0.18258	0.12576	0.08662	0.05966
Life Expectancy at Age 30									38.9 years	43.2 years	47.8 years	53.1 years	59.7 years

Note: The percentages in the table represent the education quintiles (20% groupings). Life expectancy at age 30 was computed using standard life table techniques and graduation to compute $_na_x$ values.

References:

Bound, J., Geronimus, A. T., Rodriguez, J. M., & Waidmann, T. A. (2015). Measuring Recent Apparent Declines in Longevity: The Role Of Increasing Educational Attainment. *Health Affairs*, *34*(12), 2167–2173. doi:10.1377/hlthaff.2015.0481

Dowd, J. B., & Hamoudi, A. (2014). Is life expectancy really falling for groups of low socio-economic status? Lagged selection bias and artefactual trends in mortality. *International Journal of Epidemiology*, *43*(4), 983–988.

Hendi, A. S. (2015). Trends in U.S. life expectancy gradients: the role of changing educational composition. *International Journal of Epidemiology*, *44*(3), 946–955. doi:10.1093/ije/dyv062

Hendi, A. S. (2017). Trends in Education-Specific Life Expectancy, Data Quality, and Shifting Education Distributions: A Note on Recent Research. *Demography*, *53*(3), 1203–1213.