[FOR ONLINE PUBLICATION]
Web Appendix to "Understanding the Mechanisms Linking College Education with Longevity"

Kai Hong[1] Peter Savelyev[2] Kegon T. K. Tan[3]

New York University College of William & Mary University of Rochester

April 2, 2020

[1]Robert F. Wagner Graduate School of Public Service, New York University. Address: 295 Lafayette St, New York, NY 10012. E-mail: kai.hong@nyu.edu.

[2]The corresponding author. Department of Economics at The College of William & Mary and IZA. Address: 300 James Blair Drive, Tyler Hall Room 317, Williamsburg, VA 23185. E-mail: pasavelyev@wm.edu.

[3]Department of Economics, The University of Rochester. Address: 280 Hutchinson Road, P.O. Box 270156, Rochester, NY 14627. E-mail: ttan8@ur.rochester.edu.

Contents

A Supplementary Figures and Tables

This appendix supplements the main paper with additional tables and figures, which are listed in the order they are mentioned in the main paper. Below we provide short summaries of these supplemental results.

Robustness with Respect to the Method of Measurement of Latent Skills It is a potential weakness of the full maximum likelihood approach that, in the case of a misspecified model, a full maximum likelihood method could lead to a misleading measurement of latent factors: latent skills might be primarily measured not by specific early measures of skills but by outcomes (e.g., Heckman et al., 2016, 2013). For instance, in this paper, latent skills might be mainly measured not by school achievement but by outcomes such as the hazard of death and smoking. We show that this is not the case for our model.

Following Heckman et al. (2016), we perform an alternative estimation. First, we estimate the measurement system for skills. Then, we proceed with full maximum likelihood estimation, but fix the factor loadings in the measurement system to numbers estimated in the first step, thus preventing them from being biased by outcomes in the case of possible model misspecification.

By comparing the solid and hollow rounds in Figure A-1 representing the two alternative approaches, we can see that results are virtually identical. We conclude that our results do not suffer from a bias due to inclusion of outcomes to the factor model. We therefore use the full likelihood estimation, which is a more efficient method than its two-step alternative. These results are consistent with correct specification of our model.

Testing for Links between Key Variables and Survival to 1993 In Table A-1, we use a similar model as used for the main paper, but test whether there is no conditional correlation between surviving from 1957 to 1993 and key skill variables: latent skills, IQ, and education. The table shows single- and joint test p-values, which range from 0.36 to 0.96, suggesting that we cannot reject the null hypothesis.

Comparison of the Main Model Results with a Model that Omits Behaviors in the Health Stock Equation, A Robustness Check Tables A-2 and A-3 compare two versions of the model: (1) the main model (see Equations (1–8)) and its version with the only difference being that in Equation (3) there is no control for behaviors, B_k. The formula for decomposition (9) is adjusted accordingly: coefficients c_{4k} are set to zero. We can see that results are robust to this change in model specification: for instance, the difference in the estimate of total effect for men (-0.32 vs. -0.33) is negligible given the size of the standard error (st. err. =0.01), which is an order of magnitude higher than the difference, 0.1 (see Table A-2). For women, despite high standard errors, the estimate of the total effects for women is the same in both cases, -0.080 (if using 3 decimal places, see Table A-3).

In the main model, estimates for behaviors tend to be marginally larger than in the robustness check and estimates for health stock tend to be marginally smaller, but these differences are so small than they hardly make a difference.

The likely reason for robustness of the decomposition is that in the absence of behaviors in Equation (3) we simply have a reduced form model for health, while the link between behaviors and longevity is still captured by Equation (4).

Choosing the Optimal Number of Latent Classes Table A-4 documents AIC and BIC of the main model as a function of the number of latent classes. Following Nylund et al. (2007), we determine that the optimal number of latent classes $q^{max} = 3$ for both men and women, as $q^{max} = 3$ minimizes both AIC and BIC.

Placebo tests Table A-5 documents placebo tests for the main model. Placebo tests use early health-related outcomes that should not be affected by education, as if they were late life outcomes that may be affected. Here we replace the duration model for the hazard of death with a logit model for a placebo outcome. The placebo test is passed if we do not reject the null hypothesis of no effect of education on the placebo outcome.

Ideally, it would be perfect to perform placebo tests on measures of early health that are not a part of the model, but we do not have such measures. The study started at the end of high school and had a different focus than measuring early health. Therefore, we do the best that we can given the data by (1) using proxies of health rather than true early health measures; (2) taking these proxies from the set of background controls $\boldsymbol{X}$ of the model one by one, not from data unused by the model. As the procedure involves testing multiple hypotheses, we adjust hypotheses rejection thresholds to control for familywise error rate (e.g., Westfall and Young, 1993).

We use information on birth order, as the age of the mother at birth and amount of resources available per child differ by birth order and may affect health. We also use information about being underweight and overweight, as well as having a smoker in the household, which is a proxy for passive smoking by the child.

Even though the procedure of taking placebo outcomes one by one from the set of background controls creates a risk that hypotheses get rejected not because of poor model specification but because we no longer use the full set of original controls $\boldsymbol{X}$, we cannot reject the null.

Estimates of Factor Loadings Table A-6 shows estimates of factor loadings from the measurement system. All factor loadings are of comparable magnitude, are statistically significant at the 1% level, and have signs expected from theoretical considerations.

Comparison with Models that Omit Essential Controls Table A-7 presents the average bias in decompositions generated by the failure to account for one or more types of controls including (1) unobserved heterogeneity, $\boldsymbol{\mu}$; (2) latent skills, $\boldsymbol{\Theta}^S$; and (3) traditional observed controls, $\boldsymbol{X}$.

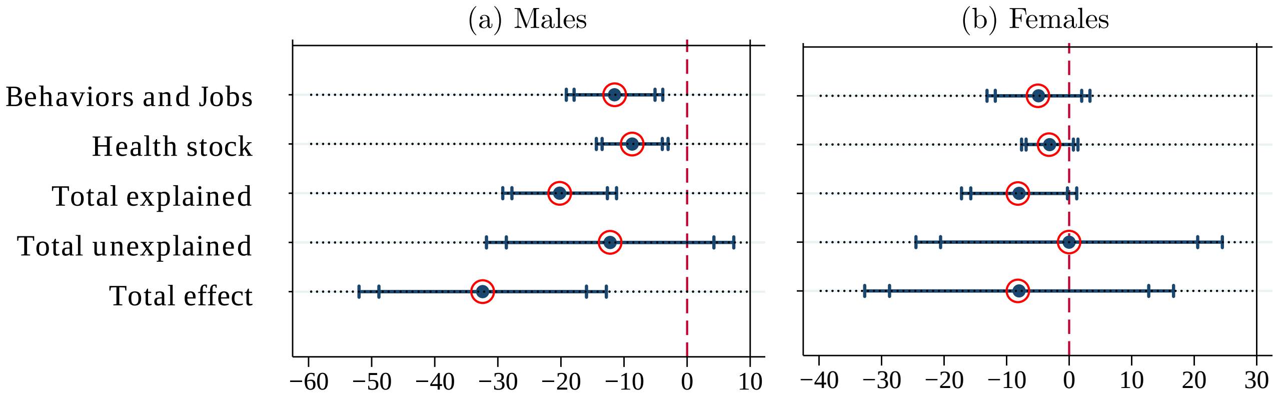
Omitting the control for unobserved heterogeneity leads to an 8–31% bias. Further omitting latent skills leads to a 8–35% bias relative to results conditional on the full set of controls. Finally, omitting all controls leads to a 37–213% bias. Therefore, results of the model differ greatly from unconditional results (up to more than 200%). In addition, omitting latent skills and unobserved heterogeneity makes a sizable numerical difference for our results (up to 35%) despite using a detailed set of observable controls, $\boldsymbol{X}$.

Estimates of the Analogue of the Main Model with Health Stock Used as the Final Outcome Table (A-8) presents decompositions that are based on a modification of the main model (1–8), in which Equation (4) for the hazard of death between years 1992 and 2017 is replaced with a similar equation for health stock in year 2011 as an outcome.

For health stock in year 2011 we use a standardized factor score. The factor model is based on the following measures of health in 2011: (1) general health, (2) having a major illnesses (3) stayed in bed at least once last year. These are the same variables as we use for 1992 health stock (see variable definitions in Table 2 of the main paper), except for "hospitalization at least once last year," which is not available for 2011.

The results show a different margin of health production than the main model does. Like in the results for longevity, we see the total effect of education on health for men (about 0.19 standard deviation increase in standard deviation of latent health). We also find a statistically significant total effect of education on health for women (about 12% of a standard deviation increase). For men, health stock is a statistically significant component. Other estimated contributions are in the expected direction, but they are not precisely determined. For women we see contributions from exercise, being overweight, and income, while health stock is not precisely determined.

Figure A-1: Explained, Unexplained, and Total Effects of Education on the Hazard of Death, a Comparison of the Main Model with a Model with Factor Loadings in the Measurement System Estimated Using Early Measures Only, a Robustness Check, %



Notes: Hollow rounds represent a robustness check when loadings of the measurement system for skills in the full ML model are fixed at the level determined by the measurement system estimated separately from outcomes as the first step. Panels (a) and (b) share a common scale and represent decomposition (9) for $\Delta = 1$. Inner and outer vertical bars represent the 90% and 95% Huber-White confidence intervals. Calculations are based on the WLS data.

Table A-1: Testing for Links between Key Skill Variables and Survival to 1993: Single and Joint Tests' p-Values

	Males	Females
Latent Skills	0.963	0.472
IQ	0.467	0.731
College Degree	0.645	0.873
Joint Test (Wald)	0.5367	0.3605
Estimation Sample Size	4152	4650

Notes: p-values for individual t-tests and joint Wald tests are shown.

Table A-2: Comparison of the Main Model Results with the Model that Omits Behaviors in the Health Stock Equation, A Robustness Check, Males

	Main model (of Mortality)[a]			No Behaviors in Health Eqn[b]		
	Estimate	Std. err.	p-value	Estimate	Std. err.	p-value
Aggregated components				starts 0		
Total Behaviors and Job	-0.115 ***	0.039	0.003	-0.107 **	0.045	0.018
Total Explained	-0.202 ***	0.046	0.000	-0.199 ***	0.062	0.001
Total unexplained	-0.122	0.100	0.225	-0.133 ***	0.104	0.200
Total[c]	-0.324 ***	0.100	0.001	-0.333 ***	0.097	0.001
Specific components						
Smoking	-0.013 *	0.007	0.055	-0.014 **	0.007	0.053
Risky Drinking	-0.005	0.005	0.326	-0.006	0.005	0.277
Exercise	-0.008	0.005	0.124	-0.006	0.004	0.185
Overweight	-0.017 **	0.009	0.046	-0.012 *	0.007	0.092
Married	0.006	0.038	0.866	-0.002	0.009	0.850
Social Activity	-0.005	0.008	0.494	-0.005	0.007	0.511
Income	-0.055	0.034	0.104	-0.045	0.033	0.175
Dangerous Job	-0.018 *	0.010	0.063	-0.018 *	0.010	0.063
Health Stock[c]	-0.087 ***	0.029	0.002	-0.092 ***	0.034	0.003
Sample size		3961			3961	

Notes: [a]Decomposoitions of the hazard of death based on Equation (9). Huber-White standard errors are shown. Asterisks denote the level of statistical significance: ***, **, and * denote $p < 0.01$, 0.05, and 0.10. [b]Decomposoitions of the hazard of death are based on an Equation similar to (9) but adjusted for no effect though behaviors affecting health stock. [c]One-sided test for this total effect is motivated by abundant evidence from the literature that the total effect of education on longevity is nonnegative (Grossman, 2006; Grossman and Kaestner, 1997; Lochner, 2011).

Table A-3: Comparison of the Main Model Results with the Model that Omits Behaviors in the Health Stock Equation, A Robustness Check, Females

	Main model (of Mortality)[a]			No Behaviors in Health Eqn[b]		
	Estimate	Std. err.	*p* -value	Estimate	Std. err.	*p* -value
Aggregated components				starts 0		
Total Behaviors and Job	-0.049	0.042	0.241	-0.044	0.040	0.270
Total Explained	-0.080 *	0.047	0.089	-0.092 *	0.050	0.065
Total unexplained	0.000	0.125	0.998	0.012	0.125	0.926
Total[c]	-0.080	0.126	0.264	-0.080	0.127	0.264
Specific components						
Smoking	-0.009	0.008	0.246	-0.010	0.008	0.237
Risky Drinking	0.000	0.001	0.934	0.000	0.001	0.765
Exercise	-0.015 *	0.008	0.069	-0.010	0.007	0.162
Overweight	-0.011	0.007	0.110	-0.007	0.005	0.187
Married	0.033	0.022	0.123	0.029	0.019	0.143
Social Activity	0.004	0.019	0.821	0.003	0.019	0.878
Income	-0.050 **	0.024	0.039	-0.047 *	0.024	0.051
Dangerous Job	-0.001	0.002	0.585	-0.001	0.002	0.637
Health Stock[c]	-0.031 *	0.023	0.089	-0.048 *	0.025	0.029
Sample size		4491			4491	

Notes: [(a)]Decomposoitions of the hazard of death based on Equation (9). Huber-White standard errors are shown. Asterisks denote the level of statistical significance: ***, **, and * denote $p < 0.01$, 0.05, and 0.10. [(b)]Decomposoitions of the hazard of death are based on an Equation similar to (9) but adjusted for no effect through behaviors affecting health stock. [(c)]One-sided test for this total effect is motivated by abundant evidence from the literature that the total effect of education on longevity is nonnegative (Grossman, 2006; Grossman and Kaestner, 1997; Lochner, 2011).

Table A-4: Choosing the Optimal Number of Latent Classes by Minimizing AIC and BIC

latent classes, q^{max}	males		females	
	AIC	BIC	AIC	BIC
no latent class	85154.05	87711.74	94041.51	96650.31
2	84368.21	87045.30	93223.14	95953.72
3	**84057.02**	**86840.94**	**92582.31**	**95421.87**
4	84089.69	86923.88	92598.36	95489.20
5	84113.57	86998.04	92614.37	95556.48

Notes: The minimal number in each column is in bold. Results are based on our main model applied to the WLS data.

Table A-5: Total Effects of Education on Placebo Outcomes and Placebo Tests

Variable	Unadjusted results total effect	std. error	unadjusted *p*-value	Bonferroni thresholds 5%	10%	Holm-Bonferroni thresholds 5%	10%	Result
Males								
First-born or the only child	0.023	0.024	0.324	0.007	0.014	0.008	0.017	stat. insignificant
Second-born	-0.023	0.021	0.273	0.007	0.014	0.007	0.014	stat. insignificant
Third-born	-0.013	0.126	0.916	0.007	0.014	0.050	0.100	stat. insignificant
Fourth-born or above	0.004	0.016	0.813	0.007	0.014	0.017	0.033	stat. insignificant
Respondent is overweight[a]	0.013	0.021	0.539	0.007	0.014	0.013	0.025	stat. insignificant
Respondent is underweight[b]	-0.019	0.148	0.899	0.007	0.014	0.025	0.050	stat. insignificant
Childhood household had a smoker[c]	-0.021	0.028	0.463	0.007	0.014	0.010	0.020	stat. insignificant
Females								
First-born or the only child	-0.038	0.025	0.120	0.007	0.014	0.008	0.017	stat. insignificant
Second-born	0.026	0.022	0.235	0.007	0.014	0.010	0.020	stat. insignificant
Third-born	0.006	0.017	0.739	0.007	0.014	0.013	0.025	stat. insignificant
Fourth-born or above	-0.001	0.018	0.941	0.007	0.014	0.050	0.100	stat. insignificant
Respondent is overweight[a]	-0.048	0.022	0.028	0.007	0.014	0.007	0.014	stat. insignificant
Respondent is underweight[b]	0.002	0.014	0.879	0.007	0.014	0.017	0.033	stat. insignificant
Childhood household had a smoker[c]	0.003	0.023	0.904	0.007	0.014	0.025	0.050	stat. insignificant

Notes: The table calculates the total effect according to our main model using a number of proxies of early health instead of longevity (duration model is replaced with a logit model). Health-related controls are removed one-by-one from $\boldsymbol{X}$ and treated as if they were life outcomes in adulthood. Marginal effects are shown. Raw p-values should be compared not with standard thresholds (such as 5%), but with thresholds based on Bonferroni and Bonferroni-Holm stepdown methods to adjust for family-wise error rate (e.g., Westfall and Young, 1993). Results are based on the WLS data.

Table A-6: Estimates of the Measurement System

	Males		Females	
Achievement Measures				
Standardized academic	0.631	***	0.570	***
achievement	*(0.022)*		*(0.019)*	
Member of an honor society	1.601	***	1.534	***
	(0.154)		*(0.137)*	
Outstanding student	1.411	***	1.627	***
	(0.114)		*(0.128)*	
Health stock				
General health	0.573	***	0.562	***
	(0.037)		*(0.030)*	
Major illness	-1.229	***	-1.383	***
	(0.106)		*(0.027)*	
Stayed in bed at least once	-0.747	***	-0.691	***
last year	*(0.105)*		*(0.071)*	
Hospitalization at least	-1.244	***	-0.991	***
once last year	*(0.216)*		*(0.135)*	
Sample size	3961		4491	

Notes: This submodel of the main model is estimated based on the WLS data.

Table A-7: Average Bias Induced by Omitting Essential Controls

	Comparison 1	Comparison 2	Comparison 3
Males			
Aggregated components	7.6%	8.2%	37%
Individual mediators	5.1%	18%	114%
Females			
Aggregated components	29%	35%	213%
Individual mediators	31%	29%	100%
Missing controls that are sources of bias:			
Unobserved heterogeneity	omitted	omitted	omitted
Latent skills	controlled	omitted	omitted
Traditional controls	controlled	controlled	omitted

Notes: The table compares our main model that controls for traditional background variables, latent skills, and unobserved heterogeneity with models that lack one or several of these controls, as specified in the bottom of the table. For each estimate of the decomposition component that is statistically significant, at least at the 10% level for each alternative set of controls, the bias of alternative models is calculated in % relative to the main model counterpart. (Statistically insignificant components are excluded, as they lead to less reliable estimates of the bias.) Then, the absolute values of these biases are averaged.

Table A-8: Estimates of the Analogue of the Main Model with Health Stock (Instead of the Hazard of Death) Used as the Final Outcome

	Males			Females		
	Estimate	Std. err.	*p*-value	Estimate	Std. err.	*p*-value
Aggregated components						
Total Behaviors and Job	0.046	0.038	0.230	0.043 ***	0.017	0.009
Total Explained	0.194 ***	0.055	0.000	0.054	0.042	0.197
Total unexplained	-0.001	0.099	0.983	0.063	0.056	0.266
Total[a]	0.191 **	0.081	0.010	0.116 **	0.057	0.021
Specific components						
Smoking	0.001	0.003	0.552	0.002	0.003	0.593
Risky Drinking	0.003	0.004	0.448	-0.002	0.002	0.279
Exercise	0.007	0.006	0.271	0.011 *	0.006	0.075
Overweight	0.012	0.022	0.602	0.019 **	0.009	0.039
Married	0.007	0.015	0.609	0.000	0.000	0.727
Social Activity	-0.001	0.003	0.814	0.003	0.009	0.719
Income	0.007	0.018	0.670	0.011 *	0.006	0.082
Dangerous Job	0.007	0.007	0.312	0.000	0.001	0.861
Health Stock[a]	0.148 ***	0.056	0.004	0.010	0.036	0.389
Sample size		2413			2844	

Notes: Decompositions of the effect of college education on health in 2011 (about age 72) based on Equation (9). Huber-White standard errors are shown. Asterisks denote the level of statistical significance: ***, **, and * denote $p < 0.01$, 0.05, and 0.10. [a] The one-sided test for this total effect is motivated by abundant evidence from the literature that the total effect of education on longevity is nonnegative (Grossman, 2006; Grossman and Kaestner, 1997; Lochner, 2011).

B A More General Model Specification

Here we argue that it is unnecessary to further complicate our main parsimonious model (1–8) with additional degrees of freedom. The main motivation for the parsimonious model is practical: additional degrees of freedom lead to either a model that takes an impractically long time to perform estimation and specification tests or, in the worst cases, does not lead to numerically stable estimates (lack of empirical identification). Additionally, elimination of redundant degrees of freedom should, theoretically, improve the efficiency of estimators.

In the absence of any established theoretical predictions regarding nonlinearities of this model, whether adding nonlinear terms is beneficial becomes an empirical question. Consider a model that has additional interactions between IQ and skills, education and skills, mediators and skills, and mediators and education. Moreover, potentially endogenous early-life skills, Θ^S, and early-life behaviors, B^*_{0p}, can be more explicitly linked to unobserved heterogeneity μ (see Equations (B.1) and (B.2) below). Following notation that is similar to the main text, we can write a more general recursive submodel as follows (all estimation is conditional on X, which we do not show for shortness):

$$\Theta^S = \mu_S + \epsilon_S \tag{B.1}$$

$$B^*_{0p} = a_{1p}\Theta^S + a_{2p}\Theta^S \cdot IQ + \mu_{B0p} + \epsilon_{B0p},\ p = 1, ..., Q \tag{B.2}$$

$$D^* = b_1\Theta^S + b_2\Theta^S \cdot IQ + \sum_p b_{3p}B_{0p} + \mu_D + \epsilon_D \tag{B.3}$$

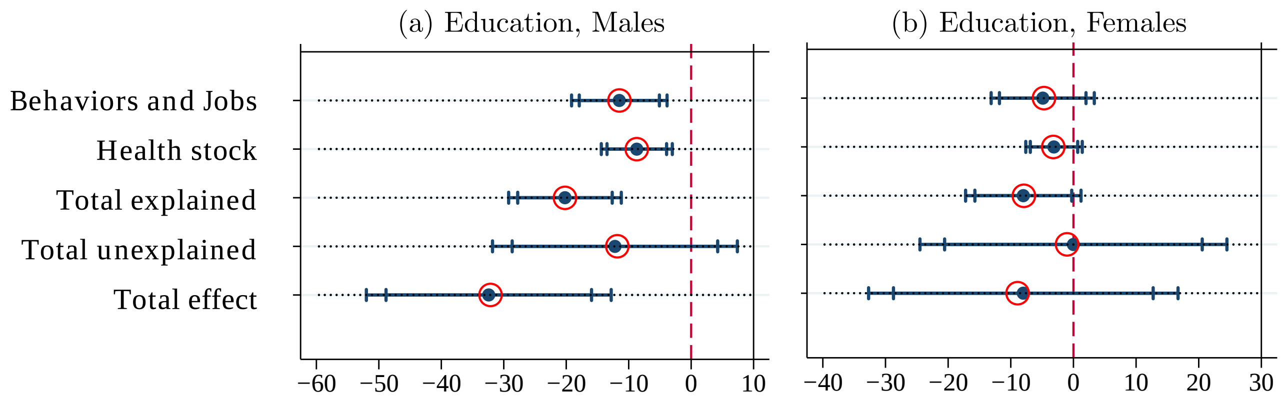
$$B^*_{1k} = c_{1k}\Theta^S + \sum_p c_{2kp}B_{0p} + c_{3k}D + c_{4k}\Theta^S \cdot IQ + c_{5k}\Theta^S D + c_{6k}IQ \cdot D + \mu_{B1k} + \epsilon_{B1k},$$
$$k = 1, ..., K \tag{B.4}$$

$$\Theta^H_1 = d_1\Theta^S + \sum_p d_{2p}B_{0p} + d_3D + d_4\Theta^S \cdot IQ + d_5\Theta^S D + d_6IQ \cdot D + \mu_{H1} + \epsilon_{H1}, \tag{B.5}$$

$$\ln(\lambda(t)) = e_{1j}\Theta^S + \sum_p e_{2p}B_{0p} + e_{3j}D + \sum_k(e_{4k}B_{1k} + e_{5k}B_{1k}D + e_{6k}\Theta^S B_{1k} + e_{7k}IQ \cdot B_{1k})$$
$$+ e_8\Theta^S \cdot IQ + e_9\Theta^S D + e_{10}IQ \cdot D + e_{11}\Theta^H_1 + \mu_{\lambda j} + \ln(\lambda_0(t)),\ j = 1, ..., J. \tag{B.6}$$

As in the main text, submodel (B.1–B.6) needs measurement system (5–6) to be identified. Submodel (B.1–B.6) is so complex that we are unable to obtain a numerically stable model estimates if we use such submodel. Therefore, we test the hypothesis of our main model's sufficiency in three steps. First, we test whether adding Equations (B.1) and (B.2) to the main model makes a difference for results and find no such difference (see Figure B-1).

Figure B-1: Explained, Unexplained, and Total Effects of Education on the Hazard of Death, A Comparison with the Case When We Model Unobserved Heterogeneity as a Part of Early Skills and Health Behaviors, a Robustness Check



Notes: Hollow rounds represent a robustness check when Equations (B.1) and (B.2) are added to the main model. Panels (a) and (b) share a common scale and represent decomposition (9) for $\varDelta = 1$. Inner and outer vertical bars represent the 90% and 95% Huber-White robust confidence intervals. Calculations are based on the WLS data.

Second, we test for interactions in Equation (B.5) while keeping all other equations parsimonious as in the main model, and find no evidence for interaction terms (joint test p-values are 0.927 for men and 0.170 for women).[1] Then, we remove interactions from Equation (B.5) but allow them in all other equations of system (B.3–B.6), and test whether they are jointly zero (except for b_2, which we know to be marginally different from zero). We cannot reject this hypothesis with p-values 0.339 and 0.800 for men and women.

Informed by the proportional hazard (PH) test documented in Table C-2, we keep time-dependence of coefficients for noncognitive skills and education as in Equation (4) of the main paper, but not for the effects of IQ and mediators.

[1] Equation (B.5) is special, because it has continuous latent variables on both the right-hand and the left-hand sides.

C Determinants of Potential Mechanisms in Midlife and Effects of Mechanisms on Longevity

Here we analyze two components of our mediation analysis: (a) the effects of education on potential mediators and (b) the effects of potential mediators on the outcome of interest (longevity). These results are not intended as essential contributions, but rather as building blocks for the decompositions. In the main paper, we combine the two components to obtain a decomposition of the effects of education on longevity with respect to mediators, which is the main contribution of the paper.

C.1 Determinants of Potential Mediators in Midlife

The potential mediators at the start of the risk period in 1992 (about age 53) are health behaviors and health-related socioeconomic outcomes that we select based on prior evidence from the literature: (1) health-related behaviors, such as smoking tobacco or engaging in physical exercise (Cawley and Ruhm, 2012); (2) lifestyles, such as marriage or intensity of social life (Holt-Lunstad et al., 2010); (3) income (Cutler et al., 2011); and (4) dangerous working conditions (Viscusi, 2013). In total, we model $K = 8$ such mediators. An additional mediator is health stock at the start of the risk period. Health stock accounts for influences of mediators of longevity in earlier life.

Table C-1 investigates the effects of education on potential mediators among behavioral, socioeconomic, and health outcomes, all measured simultaneously at the start of the risk period.[2] From columns 1–9 we see that college graduates of both sexes have higher household incomes, are less likely to be overweight, and are more likely to engage in social activity and physical exercise. Additionally, educated men enjoy superior midlife health and are less likely to smoke tobacco, participate in risky drinking of alcohol, or experience dangerous working conditions.

The only exception to this pattern of potentially health-beneficial effects is marriage for women, which is negatively affected by a college degree. This result is in line with a well-

[2]See Table A-6 for estimates of the measurement system, which was estimated as a part of the same model.

Table C-1: Marginal Effects of Education on Potential Mediators at Midlife

	Smoking tobacco	Risky drinking of alcohol	Physical exercise	Overweight	Marriage	Social activity	Household income (log)	Dangerous working conditions	Health stock (direct effect)
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Males									
Bachelor's degree	-0.036 **	-0.044 ***	0.040 **	-0.057 ***	-0.014	0.063 ***	0.323 ***	-0.130 ***	0.273 ***
or above	*(0.016)*	*(0.013)*	*(0.020)*	*(0.022)*	*(0.071)*	*(0.017)*	*(0.035)*	*(0.019)*	*(0.083)*
Females									
Bachelor's degree	-0.026	-0.009	0.079 ***	-0.040 *	-0.076 **	0.216 ***	0.155 ***	-0.012	0.108
or above	*(0.020)*	*(0.010)*	*(0.025)*	*(0.021)*	*(0.034)*	*(0.028)*	*(0.051)*	*(0.017)*	*(0.078)*

Notes: Columns (1–8) show the effect of education ($\tilde{b}_{3k}$). Column (9) shows estimates of the direct effect c_3, calculated for the standardized latent health factor. Health behaviors, lifestyles, income, and work conditions are measured in 1992. Sample size is 3961 for men and 4491 for women. Huber-White standard errors are reported in parentheses. Asterisks denote the level of statistical significance: ***, **, and * represent $p < 0.01$, 0.05, and 0.10. See Table 2 of the main paper for definitions of mediators. Calculations are based on the WLS.

documented historical pattern (e.g., England and Bearak, 2012), which is largely due to past societal views regarding the role of women. Since the 1970s, the marriage gap between educated and uneducated women has closed, and possibly even has reversed sign. The change is largely related to greater marriage stability for educated women (Lefgren and McIntyre, 2006). In Section IV of the main paper we adjust our estimates for more recent cohorts that are no longer subject to this adverse influence.

Our results for the effects of college education on essential health behaviors, lifestyles, health stock, and earnings are auxiliary estimates that serve as building blocks for our main result, the decompositions. Quantitatively, these results are not directly comparable to other papers due to a combination of differences: definitions of variables, age group, population type, and the effect type (in this paper, the average treatment effect). In this respect, we complement the literature. Qualitatively, our results for the effect of college education are consistent with the literature on tobacco smoking (Buckles et al., 2016; de Walque, 2007; Heckman et al., 2016), risky drinking of alcohol (Buckles et al., 2016), income (Heckman et al., 2016), social activity (Huang et al., 2009), and health stock (Heckman et al., 2016). Our estimates for overweight status and physical exercise are not precisely determined for men, but these estimates have signs consistent with findings in the literature (Buckles et al., 2016).

C.2 Effects of Potential Mediators on Longevity

Table C-2 presents estimated parameters of the Mixed Proportional Hazard (MPH) model, showing the effects of mediators $\boldsymbol{B}$ and health stock Θ^H on the hazard of death at the start of the risk period (1992). These MPH results are not intended to be novel, but we need to have a full set of model coefficient estimates to proceed.

We can see that several behaviors show effects on mortality. As expected, smoking increases the hazard of death. Health stock decreases the hazard of death. In men, we also observe statistically significant harmful effects of risky drinking of alcohol and dangerous working conditions. Being overweight, marriage, physical exercise, social activity, income, and marriage show sizable

Table C-2: Effects of Potential Mediators on the Hazard of Death for Age 54–78 MPH Model Coefficients

	Males		Females	
	Estimates	Standard errors	Estimates	Standard errors
Smoking tobacco[(a)]	0.366 ***	0.117	0.363 ***	0.110
Risky drinking of alcohol[(a)]	0.137 *	0.123	0.068	0.131
Physical exercise[(a)]	-0.070	0.079	-0.054	0.082
Overweight[(a)]	0.115	0.096	0.071	0.098
Marriage	-0.274	0.279	-0.449 **	0.190
Social activity	-0.066	0.110	0.015	0.091
Household income[(a)]	-0.098	0.098	-0.269 ***	0.109
Dangerous working conditions[(a)]	0.133 **	0.071	0.034	0.097
Health stock[(a)]	-0.321 ***	0.052	-0.294 ***	0.053
Other controls[(b)]	Yes		Yes	
Joint test *p* -value[(c)]	0.000		0.000	
PH test *p* -value[(d)]	0.163		0.351	
Sample size	3961		4491	

Notes: Estimates of d_{4k}, $k = 1, ..., 8$, and d_5 from Equation (4) are shown. The dependent variable is log hazard of death conditional on survival to January 1993, at which time the median age is 54. Mediators are measured in year 1992. Standard errors are shown in parentheses. Asterisks denote the level of statistical significance: ***, **, and * represent $p < 0.01$, 0.05 and 0.10. Calculations are based on the WLS data. [(a)]For these mediators, asterisks correspond to one-sided tests, chosen due to evidence in the literature about the direction of the effect. [(b)]"Other controls" include latent skills, IQ, college education, early life health behaviors, and background variables. [(c)]We test and reject the hypothesis that all MPH model coefficients are jointly zero. [(d)]We test the proportional hazard (PH) assumption by allowing the MPH model coefficients to differ by age and testing whether they are the same over ages.

coefficients in the expected direction, but these estimates are not precisely determined. Note that we control for health stock at midlife, which already accumulated some effects of these behaviors in the past, thus decreasing estimates of direct effects and the chance to reject the null hypothesis.

The negative effect of marriage on mortality that we estimate is large and statistically significant for women. For men the estimate is large and negative but the effect is not precisely determined. The negative effect of marriage on the hazard of death is consistent with the meta-analysis by Manzoli et al. (2007).

Table C-2 also provides the results of proportional hazard tests. We test and do not reject the proportional hazard (PH) hypothesis that regression coefficients are constant over time for the risk period (see p-value for the PH tests in the bottom of the table). Estimates of our MPH model based on data from ages around 53 to 77 are internally valid for this period.

D Calculation and Extrapolation of the Survival Function

Based on estimated main model (1–8) we calculate the survival function $\hat{S}(t)$, which is not only useful by itself for our analysis, but is also utilized for calculations of life expectancy $\hat{e}$ and the value of remaining life $\hat{V}_R$.

Since Equation (4) is estimated in this paper for two discrete time periods $j(t) = 1, 2$, we construct the overall survival function $\hat{S}(t)$ from estimated survival functions $\hat{S}_1(t)$ and $\hat{S}_2(t)$. From the end of period 2 we use survival function $\hat{S}_3(t)$ estimated by the Centers for Disease Control and Prevention (CDC) for white men and women.[3] For a counterfactual calculation of $\hat{S}_3(t)$ by education, we blend the survival function from CDC with structure from our main model to achieve extrapolation $\hat{S}_3(t|D = d), \quad d = 0, 1$.

Each survival function $\hat{S}_j(t)$ is normalized to one at the beginning of its time period j: ($\hat{S}_j(0) = 1$ for each $j = 1, 2, 3$). We have:

$$\hat{S}(t) = \begin{cases} \hat{S}_1(t), & \text{for } 0 \leq t < t_1 \text{ (or } j = 1) \\ \hat{S}_1(t_1)\hat{S}_2(t - t_1), & \text{for } t_1 \leq t \leq t_2 \text{ (or } j = 2) \\ \hat{S}_1(t_1)\hat{S}_2(t_2 - t_1)\hat{S}_3(t - t_2) & \text{for } t_2 < t < \infty \text{ (or } j = 3)^{4}. \end{cases} \tag{D.1}$$

We calculate $\hat{S}_1(t)$ and $\hat{S}_2(t)$ using the following formula:

$$\hat{S}_j(t) = (1/N) \sum_{i=1}^{N} \sum_{q=1}^{q^{max}} \hat{p}_{qi} \hat{S}_{0jq}^{\exp(\hat{\boldsymbol{d}}_1 \boldsymbol{X}_{i1} + \hat{d}_{2j}\hat{\Theta}_i^S + \hat{d}_3 \hat{D}_i + \sum_{k=1}^{K} \hat{d}_{4k}\hat{B}_{kiq} + \hat{d}_5 \hat{\Theta}_{iq}^H + \hat{\mu}_{\lambda q})}, \quad j = 1, 2, \tag{D.2}$$

where i is an individual index; N is the estimation sample size; $\hat{p}_{qi}$ is the probability for individual i to belong to latent class q, calculated for each individual in the estimation sample

[3]For $S_3(t)$ we use the CDC survival data from age 77 to 100, see Tables 5 and 6 by Arias et al. (2019) . Between ages 100 and 105 we use an extrapolation of the CDC data, which assumes the same change in the yearly hazard rate after age 100 as the average change in the yearly rate between age 94 and 100. We set $S_3(t) = 0$ for age above 105. For conciseness, we refer to this partially constructed function $\hat{S}_3(t)$ from age 77 to infinity as the CDC data.

[4]In this paper, $t = 0$ at the start of 1993, $t_1 = 12$ at the start of 2005, and $t_2 = 24$ at the start of 2017.

using formulas (7) and (8); $\hat{\Theta}_i^S$ is a factor score imputed from factor model (5); $\hat{D}_i$, $\hat{B}_{kiq}$, and $\hat{\Theta}_{iq}^H$ are imputed using formulas (1–3); and $\hat{S}_{0jq}$ is the baseline survival function for period $j = 1,2$ and latent class $q = 1,...,q^{\max}$ estimated from the nonparametrically determined baseline hazard rate $\lambda_{0jq}(\tau)$:

$$\hat{S}_{0jq}(t) = \int_0^t \lambda_{0jq}(\tau)d\tau, \;\; 0 \le t \le t_j - t_{j-1}, \quad j = 1,2, \quad t_0 = 0. \tag{D.3}$$

A counterfactual analysis by education using the same CDC survival function $S_3(t)$ for both the educated and uneducated individuals would be a strong assumption.[5] Therefore, we offer a method of blending the CDC data with the structure and estimates of our main model.[6]

We first decompose the CDC survival curve $S_3(t)$ into a baseline survival curve $\hat{S}_{03}(t)$ and the counterfactual part for the average levels of variables:

$$\hat{S}_3(t) = \hat{S}_{03}(t)^{\exp(f(\bar{x},\ \bar{\theta}^S,\ \bar{d},\ \bar{\boldsymbol{\mu}}))}, \tag{D.4}$$

where

$$\begin{aligned} f(\bar{x},\ \bar{\theta}^S,\ \bar{d},\ \bar{\boldsymbol{\mu}}) = {} & \hat{d}_1\bar{x} + \hat{d}_{22}\bar{\theta}^S + \hat{d}_{32}\bar{d} + \sum_k \hat{d}_{4k} b_k(\bar{x}, \bar{\theta}^S, \bar{d}, \bar{\mu}_{Bk}) \\ & + \hat{d}_5 \theta^H(\bar{x}, \bar{\theta}^S, \bar{d}, \boldsymbol{b}(\bar{x}, \bar{\theta}^S, \bar{d}, \bar{\boldsymbol{\mu}}_B), \bar{\mu}_H) + \bar{\mu}_\lambda. \end{aligned} \tag{D.5}$$

Here $\bar{\mu}_{Bk}$, $\bar{\mu}_H$, and $\bar{\mu}_\lambda$ are averages of corresponding points of support over the population average likelihoods of these points of support, $\bar{p}_q$. For instance, $\bar{\mu}_\lambda = \sum_{q=1}^{q^{max}} \bar{p}_q \hat{\mu}_{\lambda q}$. Functions $b_k(\cdot)$ and $\theta^H(\cdot)$ are given by conditional expectations of formulas (2) and (3). Based on Equation

[5] CDC stratifies its survival functions by both race and sex, but we do not have access to detailed survival functions stratified by three characteristics: race, sex, and education.

[6] As with any extrapolation, it comes at a cost of assuming that model parameters are true beyond the time period for which the model is estimated. However, grounding our extrapolation on the CDC data allows us to make milder extrapolation assumptions.

(D.4) we can calculate the baseline survival function as

$$\hat{S}_{03}(t) = \exp\left(\frac{\log(\hat{S}_3(t))}{\exp(f(\bar{x},\ \bar{\theta}^S,\ \bar{d},\ \bar{\boldsymbol{\mu}}))}\right). \tag{D.6}$$

Finally, we can extrapolate the survival function conditional on education level d:

$$\hat{S}_3(t|D = d) = \hat{S}_{03}(t)^{\exp(f(\bar{x},\ \bar{\theta}^S,\ d,\ \bar{\boldsymbol{\mu}}))}, \quad d = 0, 1, \tag{D.7}$$

where $\hat{S}_{03}(t)$ is given by (D.6).

References

Arias, E., J. Xu, and K. D. Kochanek (2019). United States life tables, 2016. *National Vital Statistics Reports, Division of Vital Statistics 68*(4), 1–66.

Buckles, K., A. Hagemann, O. Malamud, M. Morrill, and A. Wozniak (2016). The effect of college education on mortality. *Journal of Health Economics 50*, 99–114.

Cawley, J. and C. J. Ruhm (2012). The economics of risky health behaviors. *Handbook of Health Economics 2*, 95–199.

Cutler, D. M., A. Lleras-Muney, and T. Vogl (2011). Socioeconomic status and health: Dimensions and mechanisms. *In S. Glied and P. Smith, eds the Oxford Handbook of Health Economics*.

de Walque, D. (2007). Does education affect smoking behaviors?: Evidence using the Vietnam draft as an instrument for college education. *Journal of Health Economics 26*(5), 877–895.

England, P. and J. Bearak (2012). Women's education and their likelihood of marriage: A historic reversal. A fact sheet prepared for the council on contemporary families. Technical report, New York University.

Grossman, M. (2006). Education and nonmarket outcomes. In E. Hanushek and F. Welch (Eds.), *Handbook of the Economics of Education*, Volume 1, Chapter 10, pp. 577–633. Amsterdam: Elsevier.

Grossman, M. and R. Kaestner (1997). Effects of education on health. In J. R. Behrman and N. Stacey (Eds.), *The Social Benefits of Education*, pp. 69–124. Ann Arbor, MI: University of Michigan Press.

Heckman, J. J., J. E. Humphries, and G. Veramendi (2016). Returns to education: The causal effects of education on earnings, health and smoking. Forthcoming, *Journal of Political Economy*.

Heckman, J. J., R. Pinto, and P. A. Savelyev (2013, October). Understanding the mechanisms through which an influential early childhood program boosted adult outcomes. *American Economic Review 103*(6), 2052–2086.

Holt-Lunstad, J., T. B. Smith, and J. B. Layton (2010, 07). Social Relationships and Mortality Risk: A Meta-analytic Review. *PLoS Med 7*(7), e1000316.

Huang, J., H. M. van den Brink, and W. Groot (2009). A meta-analysis of the effect of education on social capital. *Economics of Education Review 28*, 454–464.

Lefgren, L. and F. McIntyre (2006). The Relationship between Women's Education and Marriage Outcomes. *Journal of Labor Economics 24*, 787–830.

Lochner, L. (2011, January). Non-production benefits of education: Crime, health, and good citizenship. Working Paper 16722, National Bureau of Economic Research.

Manzoli, L., P. Villari, G. M. Pirone, and A. Bocca (2007). Marital status and mortality in the elderly: A systematic review and meta-analysis. *Social Science and Medicine 64*, 77–94.

Nylund, K. L., T. Asparouhov, and B. O. Muthen (2007). Deciding on the number of classes in latent class analysis and growth mixture modeling: A Monte Carlo simulation study. *Structural Equation Mod 14*(4), 535–569.

Viscusi, W. K. (2013, October). Estimating the value of a statistical life using census of fatal occupational injuries (CFOI) data. *Monthly Labor Review*, 1–17.

Westfall, P. H. and S. S. Young (1993). *Resampling-Based Multiple Testing: Examples and Methods for p-Value Adjustment*. New York: John Wiley & Sons Inc.