HCEO WORKING PAPER SERIES

Socioemotional Skills, Education, and Health-Related Outcomes of High-Ability Individuals

Peter Savelyev
Kegon Teng Kok Tan

Working Paper
2017-086
12/2017

HUMAN CAPITAL AND
ECONOMIC OPPORTUNITY
GLOBAL WORKING GROUP

The University of Chicago
1126 E. 59th Street Box 107
Chicago IL 60637

www.hceconomics.org

Socioemotional Skills, Education, and Health-Related Outcomes of High-Ability Individuals*

Peter Savelyev† Kegon T. K. Tan‡

The College of William & Mary The University of Rochester

November 30, 2017

*A version of this paper was presented to the 6th Annual Health Econometrics Workshop in Toronto; to the 5th ASHEcon Conference in Los-Angeles; to the IZA/OECD/World Bank Workshop on Cognitive and Noncognitive Skills in Bertinoro, Italy; to the Singapore Economic Review Conference; to the Empirical Micro Lunch at the University of Wisconsin at Madison; and to the Vanderbilt Empirical Applied Micro Work-In-Progress Lunch. We thank participants of these meetings for their productive feedback. We are also grateful to Laura Argys, Gabriella Conti, Thomas Deleire, Evan Elmore, Erik Meijer, Frank Sloan, Chris Taber, Benjamin Ward, and journal referees for their comments and suggestions, which greatly contributed to progress with this paper. Atticus Bolyard provided excellent proofreading of the manuscript. Peter Savelyev gratefully acknowledges research support from the College of William and Mary, the Grey Fund, and the ERC at the University of Chicago. Kegon Tan gratefully acknowledges the support of the Human Capital and Economic Opportunity Global Working Group sponsored by the Institute for New Economic Thinking. The authors have no potential conflicts of interest. Sponsors of this research did not participate in study design, collection, analysis, interpretation of data, or in writing the manuscript. The Terman data are provided by the ICPSR, Ann Arbor, MI. Supplementary materials may be retrieved from https://wmpeople.wm.edu/asset/index/pasavelyev/termanhealthwebappx.

†Peter Savelyev, the corresponding author, is an Assistant Professor of Economics at The College of William & Mary, Research Assistant Professor at Vanderbilt University, and Research Affiliate at IZA. Address: Economics Department, College of William & Mary, 300 James Blair Drive, Tyler Hall Room 317, Williamsburg, VA 23185, USA. Email: pasavelyev@wm.edu. Phone: 1(757)221-2371; Fax: 1(757)221-1175.

‡Kegon Tan is an Assistant Professor of Economics at the University of Rochester. Address: 280 Hutchison Road, Box 270156, University of Rochester, Rochester, NY 14627, USA. Email: ttan8@ur.rochester.edu.

Abstract

We use the high IQ Terman sample to estimate relationships between education, socioemotional skills, and health-related outcomes that include health behaviors, lifestyles, and health measures across the lifecycle. By both focusing on a high IQ sample and controlling for IQ in regression models, we mitigate ability bias due to cognitive skill. In addition, we control for detailed personality measures to account for socioemotional skills. We model skills using factor analysis to address measurement error and adopt a powerful stepdown procedure to account for multiple hypothesis testing. We find that among high IQ subjects, education is linked to better health-related outcomes, in contrast to previous evidence. Conscientiousness, Openness, Extraversion, and Neuroticism are linked to various health-related outcomes across the lifecycle. Furthermore, we find that accounting for a comprehensive set of skills, measurement error, and multiple hypothesis testing not only provides greater confidence in several established relationships but also generates novel results.

Key words: college education, Big Five personality taxonomy, health behavior, lifestyle, health

JEL codes: I12, J24

I. INTRODUCTION

There have been numerous studies on the education-health gradient, although it remains a controversial topic. Recently there has been a growing recognition of the importance of socioemotional skills for health as both major predictors of health-related outcomes and confounders of the effect of education on health. Yet it is challenging to study the roles of both education and early skills on health in mid- and late-life due to data limitations. We exploit a unique dataset, the Terman lifecycle data of children with high ability, to overcome this problem.

The Terman cohort consists of subjects born around 1910 who were selected from schools in California for their high IQs. The data prospectively cover the period from 1922 to 1991 and combine high-quality measures of IQ and personality obtained around age 12, personality around age 30, and lifecycle measurements of health and health-related outcomes. The rich and long panel data allow us to estimate relationships across multiple health-related outcomes at different stages of the lifespan.

Research on the education-health gradient has been plagued by the presence of confounding factors. Both cognitive and socioemotional skills were shown to be important sources of bias for estimates of education's effect on health (Conti and Hansman, 2013; Conti et al., 2010). In this paper, we address the potential bias issue in two ways. First, we focus on a sample of high IQ individuals, which reduces the magnitude of both the education-ability correlation and the health-ability correlation, hence lessening the potential for bias due to cognitive ability in the effect of education on health. Second, we control for an IQ score covariate in our regressions and explicitly model latent socioemotional skills that are expected to contribute to ability bias (Carneiro, Hansen, and Heckman, 2003; Heckman, Pinto, and Savelyev, 2013; Heckman, Stixrud, and Urzúa, 2006).

This paper differs methodologically from papers that rely on natural experiments as a source of identification. Given the limitations of natural experiments[1] and the controversy over the causal status of education's effect on health[2], we believe that our alternate approach provides useful evidence to complement existing studies. In particular, instruments based on compulsory schooling laws tend to recover estimates for relatively low-achieving individuals who are induced by the law to complete additional primary or secondary education, while our estimates are valid for a high-achieving population.

We acknowledge that this approach has its own limitations. Causal inference for our estimates relies on a generalization of the conditional independence assumption. We cannot rule out that some unaccounted confounder violates this assumption. However, it is difficult to consider a confounder that is not covered by our detailed set of observable and latent controls and that would be a concern across both the wide range of outcomes as well as the multiple stages of the lifecycle that we analyze. Additionally, our identification is supported by a placebo test. Further, we have evidence that using a sample of high-ability people greatly reduces the role of confounders.[3]

We find that among high IQ individuals, education is linked to multiple health-related

[1]The limitations of natural experiments include issues with validity, monotonicity, weak instruments, loss of power, and identification of the effect only for the group that is induced to change behavior by the instrument (e.g., Cameron and Trivedi, 2005; Carneiro et al., 2011; Heckman and Vytlacil, 2005, 2007).

[2]While some papers claim a causal effect of compulsory education on health or longevity (Grossman, 2004; Grossman and Kaestner, 1997; Lleras-Muney, 2005; van Kippersluis et al., 2011), others find that there is little to no effect (Albouy and Lequien, 2009; Clark and Royer, 2013; Mazumder, 2008).

[3]We find that the inclusion of IQ, skills and background controls for the Terman sample barely changes the education coefficient, confirming that using a sample selected on high IQs reduces concerns of bias (see figure C-4 of the web appendix).

behaviors and outcomes, and that these links persist after conditioning on IQ, socioemotional skills, and detailed background controls and accounting for multiple hypothesis testing. This contradicts results found by Auld and Sidhu (2005), who use parental schooling as an instrumental variable and find that for high ability individuals in the NLSY, additional schooling has little health benefit, and most of the impact of education on health is observed among low ability individuals with little schooling.[4] They conclude that policies aimed at raising college attainment among high ability individuals are unlikely to improve population health, and that ability and schooling are likely substitutes for producing health. Our findings suggest, in contrast, that higher education and IQ may not be close substitutes for producing essential health-related outcomes, so that even high ability individuals greatly benefit from completing college.

Furthermore, to the extent that our results are externally valid[5], they strengthen claims of a causal effect of college attainment on health in two ways. First, we use a more rigorous criteria for conditional associations than many other papers in the literature by controlling explicitly for multiple skills. Second, we show strong ties between college attainment and specific behaviors or lifestyles that predict mortality, shedding light on mechanisms that drive the education-health gradient.[6]

While IQ and education have long been studied in economics, socioemotional measures are still relatively new to the economics literature, although other social sciences have integrated them into their research (Borghans et al., 2008). For this study, we used personality ratings that have been linked to the Big Five taxonomy of personality to mea-

[4] Auld and Sidhu (2005) use two binary variables indicating whether health restricts (1) type of employment or (2) amount of employment. They also use a general health rating.

[5] See Section IV for considerations about external validity.

[6] See Figures C-1–C-3 in the Web appendix for survival curves by behaviors or lifestyles.

sure socioemotional skills.[7] Among many alternative socioemotional measures, the Big Five taxonomy of personality has the advantage of being rigorously derived from data and highly recognized by many researchers (John and Srivastava, 1999). In short, Big Five personality traits are the following five factors that aggregate numerous correlated facets of personality. Openness is a propensity to be intellectual and open to new experiences and ideas. Conscientiousness is a propensity to be organized, thoughtful about the future, and rule-following. Extraversion is a propensity to be energetic towards the social and material world. Agreeableness is a propensity to be pro-social. Neuroticism is a propensity to be emotionally unstable.

The Big Five is not an exhaustive representation of personality, but most other measures of personality can be mapped into some combination of the Big Five (Costa and McCrae, 1992; John, 1990). Following Borghans et al. (2008), we view the Big Five personality traits as key individual characteristics that are distinct from cognitive ability, although some of the Big Five are moderately correlated to cognition. Furthermore, the malleability of these characteristics leads us to describe them as socioemotional skills. Socioemotional skills together with cognitive ability represent the capabilities of an individual.

Socioemotional skills have been shown to be predictive of a wide variety of outcomes, including educational attainment, cognitive development, and risky behavior (see Almlund et al. (2011) for an overview). There are several theoretical hypotheses explaining why socioemotional skills may influence health behaviors and health. Skills may reflect

[7] Martin and Friedman (2000) establish strong links between Terman data measures and standard Big Five measures.

important economic preferences such as risk aversion and time preference. For instance, Conscientiousness has been shown to be negatively correlated with discount rates (Daly et al., 2009). Socioemotional skills may also act as modifiers for actions that enter the production function of health such as following complex procedures prescribed by the doctor at home. Finally, skills may affect the production of future skills which, in turn, build future health (e.g., Cunha et al. (2010)).

Using the Big Five as a representation of socioemotional skills, we find that some skills are persistently important for health outcomes while others show mixed evidence. Other studies in economics have also shown that socioemotional skills are influential in determining health outcomes, but many rely on one-dimensional representations of socioemotional skills (e.g., Cobb-Clark, Kassenboehmer, and Schurer (2014); Heckman, Humphries, and Veramendi (2014); Hong, Savelyev, and Tan (2017)). The use of one-dimensional representations is often caused by data limitations. The same limitations often lead to the use of ad hoc measures of skills. Our results thus complement the literature by breaking down the role of socioemotional skills into multiple distinct and well-established components.

The psychology literature, on the other hand, reports numerous correlations between the Big Five personality factors and health-related outcomes.[8] While informative, these papers typically focus on how a small subset of socioemotional skills (often just one) is associated with one or several health-related outcomes. Using a small subset of correlated skills in the regression does not allow for a ceteris paribus interpretation of the estimated

[8] See, e.g., Friedman (2000, 2008), Friedman et al. (1994, 1995, 1993), Hampson and Friedman (2008), and Martin et al. (2007, 2002).

relationships. The use of a small subset of outcomes raises concerns about selecting only health-related outcomes that exhibit a statistically significant change in response to skills. Also, most papers employ either a single noisy measure of each skill, or an average of several noisy measures, or factor scores. All these approaches lead to attenuation bias if applied without using bias-correction techniques (Croon, 2002; Heckman et al., 2013).

Our work thus enhances the reliability of existing literature results by applying a combination of econometric techniques that achieve four goals: (1) minimizing a potential omitted variable bias by controlling for latent socioemotional skills, IQ, and a rich set of background variables, (2) eliminating attenuation bias by explicitly modeling measurement error in measures of skills via factor analysis, (3) avoiding selective choice of outcomes that yield statistically significant effects ("cherry-picking") by considering a large set of available health-related outcomes, and (4) avoiding over-rejection of hypotheses by controlling for multiple hypothesis testing (family-wise error rate). Since we use these rigorous procedures, we expect our results to be more reliable than results in the literature. We confirm some literature results but contradict some others.

For each outcome of interest, we jointly estimate a linear-in-parameters outcome equation and a factor model that links latent skills to their multiple noisy measures. We adjust each single-hypothesis p-value to strongly control for the family-wise error rate following Romano and Wolf (2005). Table 1 aggregates qualitative results of this paper, which differ by gender.

For males, we find evidence robust to multiple hypothesis testing for linkages between skills, education, and essential life outcomes. Education and Conscientiousness are health-beneficial, while Neuroticism and Openness are health-harming. We find mixed effects for Extraversion. IQ does not play a large role within this high-IQ sample (but still shows some mixed effects). For females, we find substantial evidence for the health-harmful effect of Neuroticism. Education shows mixed effects, due to its link with a higher probability of never being married. For Openness, we find a health-beneficial effect, which contrasts with the health-harming effect that we find for men.

While finding a positive role of Conscientiousness and a negative role of Neuroticism is common in the literature, our results suggesting that Openness is linked to worse health-related outcomes in males is new. Openness is often viewed positively and is linked to improved learning (Debra A. et al., 2006), but our paper suggests that we should be cautious about fostering Openness as a skill.

[Table 1 here]

Finally, our diverse health-related outcomes over the lifecycle are potential mediators for the link between education and longevity (see Buckles et al., 2013; Savelyev, 2017).[9] We document sizeable differences in survival rates by health behaviors, lifestyles, and health status in the Web Appendix (see Figures C-1–C-3).[10] The discovered associations between education and health-related outcomes are therefore plausible mechanisms for understanding the effect of college on longevity.[11]

[9]In a companion paper based on the same data, Savelyev (2017) explores the impact of socioemotional skills and education on mortality . This paper, on the other hand, studies a large number of health-related outcomes across the lifecycle that may serve as mechanisms through which skills and education affect mortality. While closely related thematically, the two papers use different estimation strategies due to different econometric challenges related to the MPH model in the Savelyev (2017) paper vs. multiple hypothesis testing in this paper.

[10]We do not discuss earnings in this paper although they are related to health. Instead, we refer the reader to a working paper that also uses the Terman data to extensively discuss conditional associations between socioemotional skills and earnings (Gensowski, 2014). Collischon (2017) investigates the relationship between cognition, personality, and wages using the German Socio-Economic Panel.

[11]In the Terman data we lack statistical power that is needed for a complete mediation analysis, a decomposition of the effect of education on longevity with respect to mediators. We leave the longevity mediation analysis to our working paper based on the Wisconsin Longitudinal Study that has a larger sample (Hong,

II. DATA DESCRIPTION

Research presented in this paper is based on the Terman data (Terman, 1986), which prospectively follow from 1922 to 1991 a group of about 1,500 males and females who were born around 1910. The subjects were selected from public schools in California for IQs above 140.

The availability of early life personality measures in the Terman data enables the construction of latent factors that are close to the contemporary and well-established Big Five taxonomy of personality (Martin and Friedman, 2000). Our measures of Openness, Conscientiousness, and Extraversion from 1922 are observed at about age 12. Following prior work by psychologists Friedman et al. (2010, 1995, 1993), we construct our measures from 1922 as averages of teachers' and parents' continuous ratings. We supplement these data with 1940 continuous and binary measures of Agreeableness and Neuroticism self-reported at about age 30 to complete the Big Five. Table 2 lists the raw measures by factor. The measures are explained in the table note. The grouping is based on our exploratory and confirmatory factor analysis presented in the Web Appendix.[12]

[Table 2 here]

While the sample is homogenous in that the subjects are all highly intelligent, personality measures show a wide variation. In fact, there is no evidence that the subjects' measures of personality differ significantly from the general population (Friedman et al.,

Savelyev, and Tan, 2017), but describes a different population and lacks a number of variables that the Terman data have such as childhood measures of personality.

[12]See Web Appendix B for technical details of our treatment of factors. Additionally, Table C-1 of the Web Appendix documents correlations among the Big Five factors, which are similar to those found in the literature, as we argue in Section C of the Web Appendix. This similarity provides additional evidence of the quality of our measures of personality.

1993; Terman and Sears, 2002), with a possible exception of Openness (DeYoung et al., 2005). The Terman study has an attrition rate of less than 10%, which is low for a 70-year-long prospective study.

The wealth of information in the Terman data is remarkable. Some 4,500 measurements include detailed family background, parental investments in children, personality, and health measures in childhood and adolescence, among other important determinants of health behavior and educational attainment of the subjects. Table 3 presents health-related outcomes that we explore in this paper including health behaviors and their proxies, lifestyles, and general heath measures. Many of these outcomes were observed at multiple points over the lifecycle. Table 4 describes education, IQ, and background variables.

[Tables 3 and 4 here]

In our study we excluded subjects who were not born in the period 1904–1915 (to cut out the small number of respondents born too far before or after the main cohort), subjects who have no personality ratings in 1922, subjects who are high-school dropouts, subjects with serious diseases in their early life, such as chorea or Hodgkin's disease, and subjects without information on educational attainment. Aside from excluding subjects due to missing data, these restrictions remove outliers[13] and help minimize possible reverse causality between education and health.[14]

[13] For example, 16 subjects who were high school dropouts despite extraordinary IQ.

[14] For example, subjects with serious early health problems that may have severely affected their schooling choice. Due to resampling methods that we use, controlling for binary variables that control for conditions of only a few people is not the best option due to the risk of a perfect collinearity problem in a number of resampling draws.

III. METHODOLOGY

We use a linear model to examine the associations between college education, socioemotional skills (conditional on education) and health related outcomes. Socioemotional skills may impact outcomes directly or indirectly through education. However, for the Terman data, we show that the direct channel is indistinguishable from the total impact, so that controlling for education does not lead to an underestimate of the importance of skills.[15] Let H^k be the kth health-related outcome available in the Terman data, $k \in \{1, ..., K\}$. We estimate a system of equations:

$$H^k = a^k D + \boldsymbol{b}^k \boldsymbol{\Theta}^{SE} + c^k \Theta^C + \boldsymbol{d}^k \boldsymbol{X} + \epsilon^k \tag{1}$$

$$\boldsymbol{M}^* = \boldsymbol{\psi} \boldsymbol{\Theta}^{SE} + \pi A + \boldsymbol{\gamma} \boldsymbol{X} + \boldsymbol{\eta}, \tag{2}$$

where D is an indicator for college completion; $\boldsymbol{\Theta}^{SE}$ is a vector containing latent socioemotional skills; Θ^C represents cognitive skills as measured by IQ;[16] $\boldsymbol{X}$ is a vector of background variables (see Table 4); $\boldsymbol{\eta}$ and ϵ^k are mutually independent i.i.d. error terms; $\boldsymbol{M}^*$ is a vector of both observed and latent personality measures;[17] A is the age at which personality was measured; and ψ is a matrix of factor loadings. Without loss of

[15] See Tables C-2 and C-3 of the Web Appendix that show a negligible difference between estimated direct and indirect effects of skills. The same table shows that the education coefficient is robust to controlling for only 1922 measures of skills vs. both 1922 and 1940 measures.

[16] We only have one measure of IQ, so we cannot include intelligence in the factor model due to data limitations. Since this is a study of high-IQ people with diverse socioemotional skills, the effect of IQ is of secondary importance (we can expect the effect of IQ to be, at best, weak.).

[17] Model (2) is written in a way that saves notation. Some elements of vector $\boldsymbol{M}^*$ are continuous, for which we use a linear factor model; some others are binary, for which we use a logit model representation (see Table 2 for the description of personality measures). For a continuous measure j, the asterisk can be omitted: $M_j^* = M_j$. For a latent measure i, latent M_i^* translates to $M_i = 1$ if $M_i^* \geq 0$ and to $M_i = 0$ otherwise.

generality, the variance of each element of $\boldsymbol{\Theta}^{SE}$ is set to one.[18] No restriction on correlations among latent factors is imposed. Finally, without loss of generality, signs of factor loadings are set such that the Big Five traits can be interpreted in the traditional way.[19] Identification of such factor models is standard (e.g., Anderson and Rubin, 1956).[20] We estimate system (1-2) for each $k \in \{1, ..., K\}$, allowing us to identify the effect of latent factor $\boldsymbol{\Theta}^{SE}$ on H^k whilst accounting for measurement error in measures, which is explicitly modeled in (2).[21] We estimate these systems of equations using the maximum likelihood approach.

We further estimate a restricted model that omits latent socioemotional skills

$$H^k = a_r^k D + c_r^k \Theta^C + \boldsymbol{d}_r^k \boldsymbol{X} + \epsilon_r^k, \tag{3}$$

and compare the coefficient of determination (R^2) of models (1) and (3) for each k.[22]

1. Multiple-Hypothesis Testing Problem and the Stepdown Procedure A major challenge in exploring treatment effects on multiple outcomes is accounting for false rejections due to the multiplicity of single hypotheses being tested (e.g., Westfall and Young, 1993). It

[18]This is a standard normalization for a latent factor, which has no natural metric, needed for identification of the factor model.

[19]With this standard setting needed for identification, each Big Five factor can be interpreted as implied by its names. For instance, higher Conscientiousness implies higher propensity to be organized etc., not lower. Higher Neuroticism implies lower emotional stability.

[20]Web Appendix B contains technical details about the measurement system (2), specifically about restrictions on matrix ψ that make the factors interpretable as the Big Five.

[21]Theoretically, it might be beneficial to estimate equations (1) for all k simultaneously, but this approach leads to a complex model with too many degrees of freedom, which is difficult to reliably estimate in practice. Plus, estimation of outcomes one-by-one can be expected to be more robust to misspecification. An occasional misspecification of equation (1) for one particular outcome does not bias results for all other outcomes through common factor $\boldsymbol{\Theta}^{SE}$.

[22]For equation (1), we determine R_k^2 as $1 - \widehat{Var}(\epsilon^k)/\widehat{Var}(H^k)$; for equation (3) $R_{kr}^2 = 1 - \widehat{Var}(\epsilon_r^k)/\widehat{Var}(H_r^k)$.

is well known that as the number of single hypotheses under consideration increases, the probability that at least one of them is falsely rejected given that all of them are true quickly increases. While this problem is well-recognized in genetics research, where thousands of single hypotheses are tested, in the economics literature it is largely neglected despite substantial probabilities of false rejection.

Consider a family of single tests. Let the chance of false rejection for each individual test be $\alpha = P(H_1|H_0)$. Define the family-wise error rate, FWE= $P($Reject at least one $H_i|$ all H_i are true$)$. For instance, let α be 0.05 for each single test. Then, for a family of four independent tests, the FWE(4)=$1-(1-0.05)^4$=0.19. Likewise, FWE(7)=0.30; FWE(10)=0.40; FWE(60)=0.95; and FWE(90)=0.99. Therefore, risks of false rejection are unacceptably high even for small families of tests and need to be controlled for.

The particular grouping of hypotheses into families for which p-values get adjusted is up to the econometrician. It is straightforward to consider a family that contains all single hypotheses that are tested in this paper, but this approach is overly-conservative, leading to the opposite problem: after such an adjustment, we risk accepting most, if not all, of individual hypotheses that are in fact false.

We can improve statistical power by using *a priori* information and by asking more precise research questions. Following Heckman et al. (2010), we account for multiple-hypothesis testing within each group of single hypotheses that are clustered *a priori* by type of outcome. In doing so, we account for the multiplicity of similar outcomes. For example, based on prior research, we expect that education negatively affects heavy drinking, while Extraversion has the opposite effect, but we are less sure of the age at which we should expect this effect for our population (Conti and Hansman, 2013;

Cookson, 1994; Crum et al., 1993; Droomers et al., 1999; Flory et al., 2002). We therefore form a family of heavy drinking variables measured at various ages. In addition, we consider the following groupings: (1) all available lifecycle-aggregated outcomes together and (2) all available midlife outcomes together. This way, we account for the multiplicity of hypotheses on diverse health-related outcomes (1) over the lifecycle and (2) at midlife, by which time subjects should have substantial variation in both health and addiction capital stocks.

To account for the multiple hypothesis testing problem, we use the stepdown algorithm (Romano and Wolf, 2005), a powerful procedure that provides adjusted p-values for each individual test. Let there be K individual hypotheses in a family of tests. Then, adaptation of the general stepdown algorithm to particular needs of this paper leads to the following procedure:

1. For each individual hypothesis in the family, obtain the true t-statistic and B bootstrap t-statistics. (Use absolute values of t-statistics since all tests are two-tailed.)

2. Find the maximal t-statistic among K true t-statistics. Do the same for each pseudo-sample to get a bootstrap distribution of maximal t-statistics.

3. Use the distribution of maximal bootstrap t-statistics to test the hypothesis associated with the maximal true t-statistic. The p-value of this test is the stepdown-adjusted individual hypothesis p-value for the hypothesis associated with the maximal true t-statistic.

4. Exclude the hypothesis tested in step 3 from the family for further steps. If only one hypothesis is left after the exclusion then test this hypothesis individually and stop

the procedure. If multiple hypotheses are left then repeat the procedure starting from (2).

The stepdown procedure has three important advantages. First, it strongly controls for the FWE. Strong control holds regardless of which subset of hypotheses happen to be true (any partial null), while weak control holds only if all hypotheses are true (the complete null) (Westfall and Young, 1993). Second, it tests for the statistical significance of each individual hypothesis, unlike standard joint tests. Finally, it is a more powerful method than the computationally simpler Bonferroni and Holm-Bonferroni methods. Gains in power come from accounting for statistical dependencies among individual test statistics captured through resampling (Romano and Wolf, 2005). A big computational advantage of the stepdown procedure is the lack of a need to resample t-statistics again for the subsequent stages of stepdown: steps 2–4 are repeated, while step one is performed only once.

2. Assumptions and Limitations For the identification of the effect of education, we relax the traditional conditional independence assumption and replace it by its generalization: conditional on both observables and latent factors, educational choice is orthogonal to potential health-related outcomes with and without a college degree (Carneiro, Hansen, and Heckman, 2003).

The limitation of this approach is that if some important confounding factor is still not controlled for, results should be interpreted as conditional associations. Since the Terman data contain an extraordinary number of relevant observable background variables plus IQ and comprehensive personality measures, we believe that the omitted variable bias is less of a threat than with other observational studies. We conduct a placebo test that supports our conditional independence assumption. We regress predetermined health variables on education, skills, and background controls and do not reject a single test.[23] Likewise, to identify the effects of each skill, we rely on controlling for other skills and the wealth of background controls.

A conservative view of these results is to consider them as associations conditional on an extraordinarily rich set of observables and latent factors. Therefore, we use association language when we discuss estimated coefficients. Such associations are informative since they account for many important potential confounders.

IV. RESULTS AND DISCUSSION

We present a summary of our main results for health-related outcomes in Tables 5 and 6. Each cell shows the regression coefficient representing a conditional association between a skill (or education) and a health-related outcome. The association is conditional on a large set of observable background controls, latent skills and IQ. The associations are calculated for changes in skills by one standard deviation or for changes in education status from "'no completed college education" to "completed college education."

Asterisks denote the stepdown-adjusted statistical significance level within a family (or block) of outcomes of the same type marked by bold frames. Examples of such blocks include all available heavy alcohol drinking-related outcomes and all marriage-related outcomes across the lifecycle.[24] Coefficients with p-values above 0.15 are not shown in

[23]See Table C-4 in the Web Appendix.

[24]Physical exercise, BMI, and smoking are exceptions as we know them at only one specific age, and so they are not a part of a family of multiple similar outcomes observed over time.

summary tables to reduce clutter since we can hardly statistically distinguish them from zero, but all coefficients and p-values are available in the Web Appendix.[25]

The results are typeface coded so that bolded coefficients refer to associations that are considered in the literature to be beneficial for longevity (such as a decrease in heavy drinking or an increase in physical activity), and italicized coefficients refer to adverse associations. One quick way to analyze these summary tables is to study the typeface distribution over the table.

[Tables 5–6 here]

Our paper uses a methodology that directly controls for potential sources of ability bias and provides additional evidence in favor of education's effects on health-related outcomes. Quantitatively, direct comparison with the literature is difficult due to differences in the samples analyzed, as well as differences in the level of education considered. This is particularly true for studies using compulsory schooling laws as instrumental variables, since the effect is identified based on a sub-sample of students who are marginal dropouts and focuses on the effect of another year of elementary, middle, or high school. Those affected by the schooling laws are likely to be on the lower side of the IQ distribution. In contrast, our estimates are conditional on high IQs. We therefore look to the qualitative results to facilitate comparisons with the literature.

1. Summary of Results by Type for Males Our results show that for males, college education, Conscientiousness, and marginally, Agreeableness act on health-related outcomes in a health-beneficial way and Neuroticism and Openness are disadvantageous, while

[25] See Tables A-1–A-6 of the Web Appendix for the full set of coefficients, standard errors, and p-values, both adjusted and unadjusted.

Extraversion and IQ show mixed associations (see Table 5).

Focusing on our education coefficients, we find that college is linked to an 8–11 p.p. decline in the probability of heavy drinking, a 14 p.p. reduction in the divorce rate, an 11 p.p. increase in frequent physical activity, and an 8 p.p. increase in the probability of membership to an organization. Our estimates are in line with claims in the literature that associate college with less heavy drinking (Conti and Hansman, 2013; Crum et al., 1993; Cutler and Lleras-Muney, 2010; Droomers et al., 1999), lower divorce rates (Stevenson and Wolfers, 2007), and higher physical activity (Conti and Hansman, 2013; Conti et al., 2010), but differ from other papers that find small effects of education on health behaviors (Clark and Royer, 2013). We note that the papers documenting positive education effects use years of schooling, a broader measure of education that also captures the effect of college, whereas Clark and Royer (2013) study the effect of an additional year of high school identified for marginal high school dropouts.

In the introduction, we contrasted the health-beneficial estimates of education in our high IQ sample with those found by Auld and Sidhu (2005), who find little positive effect of schooling on health for highly able individuals and therefore conclude that more schooling would have a small health impact for the highly able. This apparent contradiction might be less severe than at first glance. We do find that college does not have a robust link with our measures of general health in midlife, which is consistent with their finding that high IQ subjects with more education do not have better work-related health than high IQ subjects with less education.[26] Our disagreement stems from

[26]We find no association between education and general health of men in 1940–1960 (about ages 30–50). For women we find statistically significant effect on general health for 1940, but not for 1950–1960.

the fact that a statistically insignificant effect on general health measures in midlife does not necessarily imply a lack of impact on other health-related outcomes such as heavy drinking, physical activity, marriage, and organizational membership. In addition, in a companion paper based on the same data, Savelyev (2017) shows a strong effect of male education on longevity. We therefore dispute the view that college for high ability subjects does not have health returns.

Turning to the links between skills and outcomes, our estimates show strong and persistent health-beneficial links between Conscientiousness and heavy drinking, marriage outcomes, and mental health, consistent with the correlations reported by psychologists who worked with the Terman sample (e.g., Friedman, 2000; Goodwin and Friedman, 2006). Papers based on other data sources report similar findings (Mroczek et al., 2009; Prevoo and ter Weel, 2015). The benefits associated with Conscientiousness may stem from better self-control and better decision-making, including better matches on the marriage market, better job choices, better health investment choices, and better persistence in maintaining positive health habits.[27]

Openness exhibits negative associations with physical activity, marriage outcomes, mental health, and general health. Openness is usually viewed as a productive trait associated with being intellectual, curious, creative, and open-minded. Indeed other papers tend to find beneficial links between Openness and the outcomes that we study.[28] In this paper we condition on IQ and other Big Five personality characteristics and find associations of Openness with adverse health outcomes, unlike prior papers (e.g., Fried-

[27]See also Kern et al. (2009) for a discussion of the positive associations between Conscientiousness and health.

[28]E.g., Goodwin and Friedman (2006) find a positive link between Openness and self-reported health, while Kotov et al. (2010) find weak links between Openness and mental health problems.

man, 2000). These adverse associations could reflect the health cost of creativity, open-mindedness, and a desire for new experiences, masked by a positive correlation between Openness and IQ among other possible confounders. For example, a desire for new experiences may imply less interest in keeping a stable partner, which may explain associations with divorce-related outcomes such as "ended up divorced" and "divorced at least twice."

Neuroticism, similarly to Openness, is positively associated with divorce outcomes (borderline statistically significant coefficients), which is in line with the literature. High levels of Neuroticism are also associated with never getting married (also borderline statistically-significant). It is possibly harder for an emotionally-unstable man to find a good match or to keep relationships and family well-being on a mutually satisfactory level. For both open and neurotic men we see effects on reduced physical activity, heavy drinking, and reduced general and mental health.

Extraversion shows mixed associations. We see persistent positive associations with both heavy drinking and mental health. While the positive link with mental health is consistent with the literature (Cookson, 1994; Flory et al., 2002), the link with alcohol is in contrast to Goodwin and Friedman (2006), who find a borderline statistically significant negative link in the general US population. One possible explanation is that Extraversion can increase alcohol consumption through more participation in social events that are complementary with alcohol consumption. The increase in socializing may in turn be beneficial for mental health. Agreeableness is positively associated with membership in organizations in 1950.

Finally, associations with IQ are weak and mixed, an unsurprising result given the extraordinary intelligence of the Terman population and small variation in their IQs. We see a positive association between IQ and heavy drinking around age 30 and some mixed associations with social participation.

2. Summary of Associations by Type for Females For females, we reject a smaller fraction of hypotheses. We see that education encourages group membership among women over the lifecycle, improves their general health at least in young adulthood, and decreases their likelihood of divorce. However, the probability of never being married increases with education. The key differences with the results for males are the lack of a strong negative association with heavy drinking, a positive link with general health, and a higher likelihood of never being married.

A negative association between education and marriage for women is consistent with the literature (England and Bearak, 2012). In a decomposition based on the WLS data, Hong, Savelyev, and Tan (2017) show that the historically negative effect of education on marriage for women counterbalances a positive effect of education on wages, partially explaining the lack of a steep education-mortality gradient for women.

Similarly to males, Neuroticism is associated with lower general and mental health, but unlike males is also linked to lower BMI. To provide one possible explanation of the association of Neuroticism with lower BMI, a more neurotic female may worry more about her health or physical appearance, leading to a reduced likelihood of being overweight. The healthiness of this possible mechanism of weight reduction remains unclear. For instance, it could work through a calorically-balanced diet and active lifestyle, but it could also work through an eating disorder and excessive exercise. Cervera et al. (2003)

suggest an unhealthy channel: they find a positive association between neurotic personality and eating disorders in women 12–21 years old and note that eating disorders are much more prevalent among females.

We also see some positive associations between Extraversion and heavy drinking, the same pattern as for men. Unlike for men, the only association of Openness for women that we observe is health-beneficial: a reduction in heavy drinking.

3. Summary of Lifetime and Midlife Outcomes We also adjust inference for two alternative groupings of outcomes: (1) outcomes aggregated over the lifecycle whenever information for such aggregation is available and (2) all available outcomes at midlife. As we can see from Table 7, the key results discussed above survive this adjustment so that the qualitative summary in Table 1 is robust to this alternative approach to hypothesis grouping.[29]

[Table 7 here]

For males, as above, we see evidence that Conscientiousness and education are associated with health-beneficial outcomes, while Openness and Neuroticism are associated with health-harming ones (see Panels A and B of Table 7). Education is negatively associated with lifetime heavy drinking and divorce and positively associated with social participation. Conscientiousness is negatively associated with a higher likelihood of heavy drinking, ever smoking, mental health problems, and divorce. Neuroticism and Openness are linked to decreased mental health. In addition, Neuroticism is negatively associated with general and mental health at midlife. Extraversion is related to more

[29]Some of variables in Tables 5 and 6 are neither lifetime nor mid-life, and so the sets of available behavior types that are tested in Tables 5, 6, and 7 somewhat differ.

lifetime heavy drinking but superior mental health at midlife.

For females, as above, we see the beneficial role of education and the adverse role of Neuroticism (see Panels C and D). Education is linked to increased lifetime general health and midlife social activity and has a negative association with lifetime divorce. Neuroticism is related to diminished mental and general health.

4. Robustness to Controlling for Essential Covariates and Familywise Error Rate Research in health psychology and epidemiology has provided evidence on correlations between the Big Five factors and health-related life outcomes. However, it is unclear whether these results can survive controlling for confounding factors and multiple hypothesis testing. We demonstrate that a number of these results do survive conditioning on extraordinarily rich background characteristics, IQ, and other personality traits from the Big Five taxonomy, and correction for the family-wise error rate, while some others do not. We also show that some of these results are persistent over the lifecycle.

In particular, we verify the key role of two personality factors, Conscientiousness and Neuroticism, in influencing health (Bogg and Roberts (2004); Droomers et al. (1999); Friedman (2000); Friedman et al. (1993)). Our estimated coefficients are statistically significant and reflect a substantial percentage of sample means for many outcomes. Our results therefore confirm the positive association between Conscientiousness and health and add to a growing body of evidence suggesting that Neuroticism is a major determinant of health-related outcomes (Lahey, 2009). As mentioned, the positive association between Extraversion and drinking alcohol confirms a recognized pattern (Cookson, 1994; Flory et al., 2002).

However, some associations are less robust. Friedman et al. (1995) reports that heavy drinking is associated with Conscientiousness for both genders based on unconditional correlations, with p-values for both associations below 0.01. We find that only the association for males survives conditioning on background variables, IQ, and latent personality. For women, even before stepdown adjustment, p-values for all heavy drinking variables exceed 0.8. Similarly, in meta-analyses by Kotov et al. (2010) and Roberts et al. (2007), negative associations between Agreeableness, drinking of alcohol, and divorce are reported. We find statistically significant estimates under naive standard errors but statistical significance is lost once we control for familywise error rate.[30]

5. Predictive Power of Socioemotional Skills vs. Traditional Controls The predictive power of socioemotional skills is comparable to the combined role of education, IQ, and detailed background controls for many of the health-related outcomes of high-ability individuals. Figure 1 presents the R^2 statistic for two outcome equations: the full model (1) and the restricted model (3) that omits socioemotional skills. The results suggest that omitting socioemotional skills leads to a substantial reduction in R^2 for most of the health-related outcomes. The reduction is around 50% or higher for heavy drinking, physical exercise, and mental health. It is about 25% or higher for overweight, divorce, and general health. For smoking, the reduction is about 15% or higher. We acknowledge that in a more heterogeneous sample than Terman's, traditional controls are expected to explain a higher share of variance.

Interestingly, unlike the share of total variance explained by skills, the selection on

[30] See Web Appendix A for detailed stepdown tables.

background variables and latent skills is negligible for the Terman data. We view this not as a contradiction to previous work (Bijwaard et al., 2015; Conti et al., 2010), but as a supplementary result obtained for a high-IQ population. Figure C-4 of the Web Appendix presents education coefficients with a full set of controls and without any controls and shows that results are almost identical, with the exception of mental health.

[Figure 1 here]

6. Data Limitations While our dataset is rich in terms of the variables we observe, it has a modest sample size, which influences this paper in two main ways. First, we make linear parametric assumptions for the statistical models capturing the measurement system for socioemotional factors and the equations modeling health-related outcomes. We lack statistical power required for more complex models. Second, even though this paper suggests important health-related mechanisms through which education and socioemotional skills affect longevity, the sample is too small to allow for a full mediation analysis. Also, due to a lack of measures of Agreeableness and Neuroticism in childhood, we have to use such measures in young adulthood and interpret the resulting estimates with caution.

7. Implications of the Terman Sample and External Validity The results in this paper are based on a historical sample of people with exceptional IQs. We have access to early measures of psychological skills and high quality lifecycle data at the expense of dealing with both an unusual and deceased cohort.

Effects of education and skills may differ with the level of intelligence, so it is useful to know such effects for different levels of IQ, including the limiting case of very high IQ. This knowledge helps us verify some claims made in the literature. While Auld and

Sidhu (2005) suggest that education is only productive for health at low levels and only for low-IQ people, we find that education is highly productive at the college level even for people with extraordinarily high IQs.

Another benefit of selection on high IQ is that it reduces the potential of IQ to confound the effects of education on health. For the Terman sample we can be sure that IQs of all subjects were more than sufficient to finish college.

We do not claim applicability of the results to the general population, but the results may be applicable to a population of individuals with high IQs, though not necessarily as high as those observed in the Terman subjects. If the health benefits of skills and education depend only weakly on the exceptional IQ of our sample, then similar results may hold for less exceptional populations. The lack of interaction effects between IQ and other essential determinants of outcomes in our data supports this consideration. We are also encouraged by the fact that many of our results are consistent with results established in more general samples, making it more likely that our findings are externally valid.

Application to more recent cohorts presents another challenge. Social norms toward many of these health behaviors have changed over time. In addition, there is more information available about the effects of these behaviors on health and longevity, as well as many technological innovations that individuals can make use of to improve their health. All of these changes may affect the magnitude of the effects. That said, the qualitative results summarized in Table 1 may survive if the essential mechanisms behind the effects remain relevant. For instance, educated and conscientious people may still exhibit better self-control and decision making leading to a higher willingness and ability to engage in healthy behaviors.

V. CONCLUSIONS

The importance of socioemotional skills in the analysis of health is gaining recognition among economists. We contribute to this emerging literature by investigating the role of multi-dimensional socioemotional skills on health-related outcomes using a unique prospective lifecycle dataset with cognitive and socioemotional skills measured early in life and find substantial conditional associations after controlling for the familywise error rate.

We find that education has statistically significant conditional associations with several important health-related outcomes. This adds new evidence to the mixed results in the literature regarding the effect of education on health. We also find that the role of socioemotional skills in explaining health outcomes is comparable to that of education, IQ, and background controls combined—at least for a sample of high-IQ people. This strong result establishes socioemotional skills as an important aspect of human capital that should receive greater attention from economists.

The findings regarding socioemotional skills open additional opportunities for public policy. Conditional on the availability of socially-acceptable and cost-effective policy interventions for children, we can improve health by remediating, for instance, an extreme lack of Conscientiousness or Emotional Stability, the inverse of Neuroticism. However, relationships between Agreeableness, Extraversion, and health-related outcomes are mixed. Openness is known to be productive for a number of outcomes outside of the health domain but shows adverse association with health. Hence we are less sure that Agreeableness, Extraversion, and Openness are potentially valuable health policy targets.

References

Albouy, V. and L. Lequien (2009). Does compulsory education lower mortality? *Journal of Health Economics 28*, 155–168.

Almlund, M., A. L. Duckworth, J. J. Heckman, and T. Kautz (2011). Personality psychology and economics. In E. A. Hanushek, S. Machin, and L. Wößmann (Eds.), *Handbook of the Economics of Education*, Volume 4, Chapter 1, pp. 1–181. Amsterdam: Elsevier B. V.

Anderson, T. W. and H. Rubin (1956). Statistical inference in factor analysis. In J. Neyman (Ed.), *Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability*, Volume 5, Berkeley, CA, pp. 111–150. University of California Press.

Auld, M. C. and N. Sidhu (2005, October). Schooling, cognitive ability and health. *Health Economics 14*(10), 1019–1034.

Bijwaard, G. E., H. v. Kippersluis, and J. Veenman (2015). Education and health: The role of cognitive ability. *Journal of Health Economics 42*, 29–43.

Bogg, T. and B. W. Roberts (2004, November). Conscientiousness and health-related behaviors: A meta-analysis of the leading behavioral contributors to mortality. *Psychological Bulletin 130*(6), 887–919.

Borghans, L., A. L. Duckworth, J. J. Heckman, and B. ter Weel (2008, Fall). The economics and psychology of personality traits. *Journal of Human Resources 43*(4), 972–1059.

Buckles, K., A. Hagemann, O. Malamud, M. S. Morrill, and A. K. Wozniak (2013, July). The effect of college education on health. NBER Working Paper NO. 19222.

Cameron, A. C. and P. K. Trivedi (2005). *Microeconometrics: Methods and Applications*. New York: Cambridge University Press.

Carneiro, P., K. Hansen, and J. J. Heckman (2003, May). Estimating distributions of treatment effects with an application to the returns to schooling and measurement of the effects of uncertainty on college choice. *International Economic Review 44*(2), 361–422.

Carneiro, P., J. J. Heckman, and E. J. Vytlacil (2011, September). Estimating marginal returns to education. *American Economic Review 101*(6), 2754–2781.

Cervera, S., F. Lahortiga, M. A. Martinez-Gonzalez, P. Gula, J. de Irala-Estevez, and Y. Alfonso (2003). Neuroticism and low self-esteem as risk factors for incident eating disorders in a prospective cohort study. *Interntional Journal of Eating Disordrers 33*(3), 271–280.

Clark, D. and H. Royer (2013). The Effect of Education on Adult Mortality and Health: Evidence from Britain. *American Economic Review 103*(6), 2087–2120.

Cobb-Clark, D. A., S. C. Kassenboehmer, and S. Schurer (2014). Healthy Habits: The Connection between Diet, Exercise, and Locus of Control. *Journal of Economic Behavior and Organization 98*, 1–28.

Collischon, M. (2017, July). Returns to personality traits across the wage distribution. SOEPpapers on Multidisciplinary Panel Data Research at DIW Berlin, Working Paper.

Conti, G. and C. Hansman (2013). Personality and the education-health gradient: A note on "Understanding differences in health behaviors by education". *Journal of Health Economics 32*, 480–485.

Conti, G., J. J. Heckman, and S. Urzúa (2010, May). The education-health gradient. *American Economic Review: Papers and Proceedings 100*(2), 1–5.

Cookson, H. (1994). Personality variables associated with alcohol use in young offenders. *Personality and Individual Differences 16*, 179–182.

Costa, P. T. and R. R. McCrae (1992). Four ways five factors are basic. *Personality and Individual Difference 13*(6), 653–665.

Croon, M. A. (2002). Using predicted latent scores in general latent structure models. In G. A. Marcoulides and I. Moustaki (Eds.), *Latent Variable and Latent Structure Models*, pp. 195–223. NJ: Lawrence Erlbaum Associates, Inc.

Crum, R. M., J. E. Helzer, and J. C. Anthony (1993, June). Levels of Education and Alcohol Abuse and Dependence in Adulthood: A Further Inquiry. *Journal of Public Health 83*(6), 830–838.

Cunha, F., J. J. Heckman, and S. M. Schennach (2010, May). Estimating the technology of cognitive and noncognitive skill formation. *Econometrica 78*(3), 883–931.

Cutler, D. M. and A. Lleras-Muney (2010). Understanding Differences in Health Behaviors by Education. *Journal of Health Economics 29*, 1–28.

Daly, M., C. P. Harmon, and L. Delaney (2009, April–May). Psychological and biological foundations of time preference. *Journal of the European Economic Association 7*(2–3), 659–669.

Debra A., M., J. E. Turner, and T. D. Fletcher (2006). Linking Proactive Personality and the Big Five to Motivation to Learn and Development Activity. *Journal of Applied Psychology*.

DeYoung, C. G., J. B. Peterson, and D. M. Higgins (2005). Sources of Openness/Intellect: Cognitive and neuropsychological correlates of the fifth factor of personality. *Journal of Personality and Individual Differences 73*(4), 825–858.

Droomers, M., C. T. M. Schrijvers, K. Stronks, D. van de Mheen, and J. P. Mackenbach (1999). Educational differences in excessive alcohol consumption: The role of psychosocial and material stressors. *American Journal of Preventive Medicine 29*, 1–10.

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.

Flory, K., D. R. Lynam, R. Milich, C. G. Leukefeld, and R. Clayton (2002). The Relations Among Personality, Symptoms of Alcohol and Marijuana Abuse, and Symptoms of Comorbid Psychopathology: Results from a Community Sample. *Experimental and Clinical Psychopharmocology 10*, 425–34.

Friedman, H. S. (2000). Long-term relations of personality and health: Dynamisms, mechanisms, tropisms. *Journal of Personality 86: 6*, 1089–1107.

Friedman, H. S. (2008, July). The multiple linkages of personality and disease. *Brain, Behavior, and Immunity 22*(5), 668–675.

Friedman, H. S., P. H. Hawley, and J. S. Tucker (1994). Personality, health, and longevity. *Current Directions in Psychological Science 3*(2), 37–41.

Friedman, H. S., M. L. Kern, and C. A. Reynolds (2010). Personality and health, subjective well-being, and longevity. *Journal of Personality 78*(1), 179–216.

Friedman, H. S., J. S. Tucker, J. E. Schwartz, L. R. Martin, C. Tomlinson-Keasey, D. L. Wingard, and M. H. Criqui (1995). Childhood conscientiousness and longevity: Health behaviors and cause of death. *Journal of Personality and Social Psychology 68*(4), 696–703.

Friedman, H. S., J. S. Tucker, C. Tomlinson-Keasey, J. E. Schwartz, D. L. Wingard, and M. H. Criqui (1993). Does childhood personality predict longevity? *Journal of Personality and Social Psychology 65*(1), 176–185.

Gensowski, M. (2014). Personality, IQ, and Lifetime Earnings. IZA Discussion Paper No. 8235.

Goodwin, R. D. and H. S. Friedman (2006). Health status and the five-factor personality traits in a nationally representative sample. *Journal of Health Psychology 11(5)*, 643–654.

Grossman, M. (2004). The demand for health, 30 years later: A very personal retrospective and prospective reflection. *Journal of Health Economics 23*(4), 629–636.

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.

Hampson, S. and H. S. Friedman (2008). Personality and health: A life span perspective. In O. P. John, R. W. Robins, and L. Pervin (Eds.), *The Handbook of Personality*. New York, NY: Guilford Press.

Heckman, J. J., J. E. Humphries, and G. Veramendi (2014). Education, health and wages. Unpublished manuscript, University of Chicago, Department of Economics.

Heckman, J. J., S. H. Moon, R. Pinto, P. A. Savelyev, and A. Q. Yavitz (2010, August). Analyzing social experiments as implemented: A reexamination of the evidence from the HighScope Perry Preschool Program. *Quantitative Economics 1*(1), 1–46.

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

Heckman, J. J., J. Stixrud, and S. Urzúa (2006, July). The effects of cognitive and noncognitive abilities on labor market outcomes and social behavior. *Journal of Labor Economics 24*(3), 411–482.

Heckman, J. J. and E. J. Vytlacil (2005, May). Structural equations, treatment effects and econometric policy evaluation. *Econometrica 73*(3), 669–738.

Heckman, J. J. and E. J. Vytlacil (2007). Econometric evaluation of social programs, part II: Using the marginal treatment effect to organize alternative economic estimators to evaluate social programs, and to forecast their effects in new environments. In J. J. Heckman and E. E. Leamer (Eds.), *Handbook of Econometrics*, Volume 6B, Chapter 71, pp. 4875–5143. Amsterdam: Elsevier B. V.

Hong, K., P. Savelyev, and K. Tan (2017). Understanding the mechanisms linking personality and education with longevity. Unpublished manuscript, The College of William and Mary, Department of Economics.

John, O. P. (1990). The "big five" factor taxonomy: Dimensions of personality in the natural language and questionnaires. In L. A. Pervin and O. P. John (Eds.), *Handbook of Personality: Theory and Research*, pp. 66–100. New York: Guilford Press.

John, O. P. and S. Srivastava (1999). The Big Five trait taxonomy: History, measurement and theoretical perspectives. In L. A. Pervin and O. P. John (Eds.), *Handbook of Personality: Theory and Research* (2 ed.)., Chapter 4, pp. 102–138. New York: The Guilford Press.

Kern, M. L., H. S. Friedman, L. R. Martin, C. A. Reynolds, and G. Luong (2009). Conscientiousness, career success, and longevity. *Annals of Behavioral Medicine*. Forthcoming.

Kotov, R., W. Gamez, F. Schmidt, and D. Watson (2010). Linking "Big" Personality Traits to Anxiety, Depressive, and Substance Use Disorders: A Meta-Analysis. *Psychological Bulletin 36*(5), 768–821.

Lahey, B. B. (2009). Public Health Significance of Neuroticism. *American Psychological Association 64*, 241–256.

Lleras-Muney, A. (2005). The relationship between education and adult mortality in the United States. *Review of Economic Studies 72*(1), 189–221.

Martin, L. R. and H. S. Friedman (2000). Comparing personality scales across time: An illustrative study of validity and consistency in life-span archival data. *Journal of Personality 68*(1), 85–110.

Martin, L. R., H. S. Friedman, and J. E. Schwartz (2007). Personality and mortality risk across the life span: The importance of conscientiousness as a biopsychosocial attribute. *Health Psychology 26*(4), 428–436.

Martin, L. R., H. S. Friedman, J. S. Tucker, C. Tomlinson-Keasey, M. H. Criqui, and J. E. Schwartz (2002). A life course perspective on childhood cheerfulness and its relation to mortality risk. *Personality and Social Psychology Bulletin 28*(9), 1155–1165.

Mazumder, B. (2008). Does education improve health? A reexamination of the evidence from compulsory schooling laws. *Economic Perspectives 32*(2), 2–16.

Mroczek, D. K., A. Spiro, and N. A. Turiano (2009). Do health behaviors explain the effect of neuroticism on mortality? Longitudinal findings from the VA normative aging study. *Journal of Research in Personality 43*(4), 653–659.

Prevoo, T. and B. ter Weel (2015). The Importance of Early Conscientiousness for Socio-economic Outcomes: Evidence from the British Cohort Study. *Oxford Economic Papers 67*(4).

Roberts, B., N. Kuncel, R. Shiner, A. Caspi, and L. Goldberg (2007). The Power of Personality: The Comparative Validity of Personality Traits, Socioeconomic Status, and Cognitive Ability for Predicting Important Life Outcomes. *Perspectives on Psychological Science 2*(4).

Romano, J. P. and M. Wolf (2005, March). Exact and approximate stepdown methods for multiple hypothesis testing. *Journal of the American Statistical Association 100*(469), 94–108.

Savelyev, P. A. (2017). Socioemotional skills, education, and longevity of high-ability individuals. Unpublished manuscript, The College of William & Mary, Department of Economics.

Stevenson, B. and J. Wolfers (2007). Marriage and Divorce: Changes and their Driving Forces. *Journal of Economic Perspectives 21*, 27–52.

Terman, L. M. (1986). Terman Life-Cycle Study of Children with High Ability by Terman L. M. et al., 1922-1986 [computer file]. 2nd ICPSR release. Palo Alto, CA: Robert R. Sears [producer], 1986. Ann Arbor, MI: Inter-university Consortium for Political and Social Research [distributor], 1989. doi:10.3886/ICPSR08092.

Terman, L. M. and R. R. Sears (2002). *The Terman Life-Cycle Study of Children with High Ability, 1922–1986*, Volume 1, 1922–1928. Ann Arbor, MI: Inter-University Consortium for Political and Social Research.

van Kippersluis, H., O. O'Donnell, and E. van Doorslaer (2011). Long-run returns to education: Does schooling lead to an extended old age? *The Journal of Human Resources 46*, 695–721.

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.

Table 1: Beneficial and Adverse Effects on Health Implied by Empirical Results of This Paper, a Qualitative Summary

	Males	Females
A. Cognitive and Socioemotional skills		
Conscientiousness	+	
Openness	-	+
Extraversion	+/-	
Agreeableness		
Neuroticism	-	-[(a)]
Cognition (IQ)		
B. Formal education		
College education or above	+	+/-

Notes: "+" and "−" denote health-beneficial and adverse health effects respectively. "+/−" denote mixed health effects. Only strong evidence is summarized. [(a)]We find strong evidence of a negative effect of Neuroticism on female health, plus weak evidence of a reduction in the likelihood of being overweight. It is unclear whether this overweight reduction is healthy or unhealthy (e.g., balanced diet vs. eating disorder). Cervera et al. (2003) suggest an unhealthy channel. Calculations are based on the Terman data.

Table 2: Raw Measures Clustered by Corresponding Factors

Conscientiousness[a]	Openness[a]
Prudence	Desire to know
Conscientiousness	Originality
Truthfullness	Intelligence
Neuroticism[b]	Extraversion[a]
Miserable[c]	Fondness for large groups
Touchy[c]	Leadership
Has periods of loneliness[c]	Popularity with other children
Lonely when with others[c]	
Remorseful and regretful[c]	Agreeableness[b]
Lacks self-confidence[c]	Easy to get along with
Worrys about humiliating experiences[c]	Avoids arguments[c]
Emotionally unstable[c]	Critical[c]
Easily hurt[c]	Tactful[c]
Hard to be serene[c]	Unfeeling[c]
Moody	Domineering[c]
Sensitive	Inflated opinion of self[c]

Notes: The table shows the five factors used in this paper and corresponding measures of each factor. Measures are formulated in the questionnaires exactly as listed in this table, e.g. "prudence." Following prior work by psychologists Friedman et al. (2010, 1995, 1993), our measures from 1922 are averages of teachers' and parents' continuous ratings. Continuous and binary measures from 1940 of Agreeableness and Neuroticism are self-reported. For all continuous measures, raters are asked to put a cross on a line going from an extremely low to an extraordinary high level of the trait. The line markings provide hints on how to interpret the traits, even though names of traits are rather self-explanatory, as they are not scientific terms but normal usage words. For instance, the highest possible prudence corresponds to "always looks ahead," and "never sacrifices future good for present pleasure." The highest conscientiousness is described as "Keen sense of duty. Does right for right's sake. Always dependable." See Web Appendix B for further details about measures and measure groupings based on exploratory and confirmatory factor analysis. [a]Based on 1922 data. [b]Based on 1940 data. [c]Binary measure. Calculations are based on the Terman data.

Table 3: Health-Related Outcomes

	Males		Females	
Variable	mean	std. error	mean	std. error
Health Behaviors and proxies				
Heavy Drinking of Alcohol in 1940	0.267	(0.018)	0.102	(0.014)
Heavy Drinking of Alcohol in 1950	0.118	(0.013)	0.038	(0.009)
Heavy Drinking of Alcohol in 1960	0.347	(0.021)	0.167	(0.018)
Ever drank heavily, 1940–60	0.394	(0.019)	0.205	(0.018)
Ever smoked, 1991	0.521	(0.037)	0.425	(0.036)
Physical Excercise, 1982	0.176	(0.021)	0.173	(0.021)
Overweight BMI, 1940	0.183	(0.016)	0.063	(0.011)
Lifestyles				
Never married, 1922–86	0.061	(0.009)	0.085	(0.012)
Married once and still married, 1922–86	0.576	(0.019)	0.417	(0.022)
Ended up divorced, 1922–86	0.064	(0.009)	0.121	(0.014)
Ever divorced, 1922–86	0.267	(0.017)	0.253	(0.018)
Divorced at least twice, 1922–86	0.063	(0.009)	0.059	(0.010)
# of memberships in organizations, 1940	2.435	(0.069)	2.506	(0.083)
# of memberships in organizations, 1950	2.714	(0.090)	1.565	(0.078)
# of memberships in organizations, 1960	3.423	(0.112)	2.567	(0.124)
Ever a member of any organization, 1940–60	0.937	(0.010)	0.900	(0.014)
Health Measures				
Ever had mental difficulty, 1940–60	0.415	(0.019)	0.463	(0.022)
Never poor or fair health, 1940–60	0.926	(0.010)	0.873	(0.015)
Estimation Sample	680		527	

Notes: Calculations are based on the Terman data.

Table 4: Education, IQ, and Background Variables

Variable	Year	Males mean	Males std. error	Females mean	Females std. error
The Highest Level of Education and IQ					
Bachelor's degree or above	1922–68	0.734	(0.017)	0.686	(0.020)
IQ[a]	1922	149.3	(0.405)	148.5	(0.446)
Subject's Background					
Normal birth or no birth problems mentioned[b]	1922	0.571	(0.019)	0.629	(0.021)
No breastfeeding[b]	1922	0.091	(0.011)	0.085	(0.012)
Childhood health[c]	1922	8.526	(0.075)	9.027	(0.083)
Childhood energy[c]	1922	8.219	(0.073)	8.834	(0.078)
Participation in World War II	1945	0.410	(0.019)	-	-
Combatant in World War II	1945	0.093	(0.011)	-	-
Age in 1922	1922	11.84	(0.112)	11.80	(0.122)
Amount of private tutoring (log)[d]	1922, 28	0.105	(0.014)	0.344	(0.026)
Amount of home investment (log)[d]	1922	0.450	(0.014)	0.409	(0.016)
Cohort 1904–1907	1922	0.237	(0.016)	0.172	(0.016)
Cohort 1908–1911	1922	0.468	(0.019)	0.467	(0.022)
Cohort 1912–1915	1922	0.296	(0.018)	0.361	(0.021)
Parental Background					
Mother deceased by 1922	1922	0.028	(0.006)	0.032	(0.008)
Father deceased by 1922	1922	0.081	(0.010)	0.074	(0.011)
Parents divorced by 1922	1922	0.050	(0.008)	0.047	(0.009)
Father has at least a Bachelor's degree	1922	0.291	(0.017)	0.253	(0.019)
Parental finances are adequate	1922	0.371	(0.019)	0.384	(0.021)
Parental social standing is below average	1922	0.253	(0.017)	0.153	(0.016)
Mother is employed	1922	0.126	(0.013)	0.132	(0.015)
Father is a professional	1922	0.243	(0.016)	0.276	(0.019)
Parent born abroad	1922	0.304	(0.018)	0.267	(0.019)
Parent born in Europe	1922	0.218	(0.016)	0.202	(0.017)
Estimation Sample		680		527	

Notes: [a]The best estimate of IQ is provided by survey organizers and is based on all available information including Stanford Binet and Terman Group Tests. [b]Conditions at birth and breastfeeding were reported by parents retrospectively at the start of the study. [c]The average of teachers' and parents' ratings, each on the scale from 1 to 13. [d]Amounts of time invested from age 2 to age 7 are transformed using natural logarithm: $\ln(1 + \text{investment amount})$. Calculations are based on the Terman data.

Table 5: Summary of Effects on Health-Related Outcomes by Type, Males

	Conscientiousness	Openness	Extraversion	Agreeableness	Neuroticism	IQ	Education
A. Health behaviors and their proxies							
1940–1960 Ever Drank Heavily	**-.055** **		_.061_ **				**-.109** **
1940 Heavy Drinking	**-.046** *		_.044_			_.057_ **	**-.086**
1950 Heavy Drinking			_.040_ **		_.039_ *		**-.090** **
1960 Heavy Drinking	**-.072** **	_.056_	_.044_ *				**-.077**
1940 Overweight				**-.034**		**-.023**	
1982 Physical Activity, Freq.		_-.044_ *			_-.066_ **		**.108** *
1991 Ever Smoked	**-.107** **						
1940–1960 Any Organization							**.084** ***
1940 Number of Organizations						_-.175_ *	**.245**
1950 Number of Organizations				**.258** *			**1.172** ***
1960 Number of Organizations						**.327** **	**1.501** ***
Never Married	_.023_				_.024_		
Married Once and Still Married	**.056** *						**.120** **
Ended up Divorced	**-.023** *	_.050_ ***			_.024_		
Ever Divorced	**-.055** *						**-.137** **
Divorced at least Twice	**-.044** **	_.031_ *			_.025_		
B. Mental Health (MH)							
Ever Poor/Fair MH	**-.071** ***	_.085_ ***	**-.051** *		_.134_ ***		
1940 Mental Difficulty	**-.078** ***	_.086_ ***	**-.077** ***		_.120_ ***		
1950 Mental Difficulty	**-.040** *				_.111_ ***		
1960 Mental Difficulty	**-.080** ***	_.091_ ***	**-.101** ***		_.120_ ***		
C. General Health (GH)							
Never Poor/Fair GH		_-.032_ *			_-.021_		
1940 General Health					_-.279_ ***		
1950 General Health	**.135** **	_-.152_ **	**.096**		_-.242_ ***		
1960 General Health					_-.211_ ***		

Notes: Each cell in the table shows the regression coefficient representing a conditional association between a skill (or education) and a health-related outcome. The association is conditional on a large set of observable background controls (shown in Table 4) and latent skills. The associations correspond to changes in skills by one standard deviation (for skill coefficients) or to changes in education status from 0 to 1 (for education coefficients). Asterisks denote the stepdown-adjusted statistical significance level within a family (or block) of outcomes of the same type, such as heavy drinking outcomes by year, marked by bold frames. Coefficients are reported with accompanying statistical significance represented by asterisks, where ***,** and * indicate $p < 0.01$, 0.05, and 0.10 respectively. A coefficient with no asterisk refers to $p < 0.15$, while a blank cell refers to a coefficient with p-value above 0.15. Coefficients with p-values above 0.15 are not shown in this summary table, but are available in Tables A-1–A-6 of the Web Appendix, where we present the full set of coefficients, standard errors, and p-values, both adjusted and unadjusted. The results are typeface coded so that bolded coefficients refer to associations that are considered in the literature to be beneficial for longevity (such as a decrease in heavy drinking or an increase in physical activity), and italicized coefficients refer to adverse associations. Calculations are based on the Terman data.

Table 6: Summary of Effects on Health-Related Outcomes by Type, Females

	Conscientio-usness	Openness	Extraversion	Agreeable-ness	Neuroticism	IQ	Education
A. Health behaviors and their proxies							
1940–1960 Ever Drank Heavily		**-.073 ****	*.054* *				
1940 Heavy Drinking				**-.041 ***			
1950 Heavy Drinking							
1960 Heavy Drinking		**-.060 ***	*.049*				
1940 Overweight					**-.037 ***		**-.074 ***
1982 Physical Activity, Freq.							
1991 Ever Smoked							
1940–1960 Any Organization							**.066 ****
1940 Number of Organizations							**.789 *****
1950 Number of Organizations							**.877 *****
1960 Number of Organizations						*-.352* **	**1.213 *****
Never Married							*.074* ***
Married Once and Still Married							**.129 ***
Ended up Divorced							
Ever Divorced							**-.111 ****
Divorced at least Twice							**-.054 ***
B. Mental Health (MH)							
Ever Poor/Fair MH					*.152* ***		
1940 Mental Difficulty					*.137* ***		
1950 Mental Difficulty					*.134* ***		
1960 Mental Difficulty					*.123* ***		
C. General Health (GH)							
Never Poor/Fair GH					*-.044* ***		**.116 *****
1940 General Health				*-.133* *	*-.318* ***		**.283 ****
1950 General Health				*-.094*	*-.267* ***		**.172**
1960 General Health					*-.241* ***		

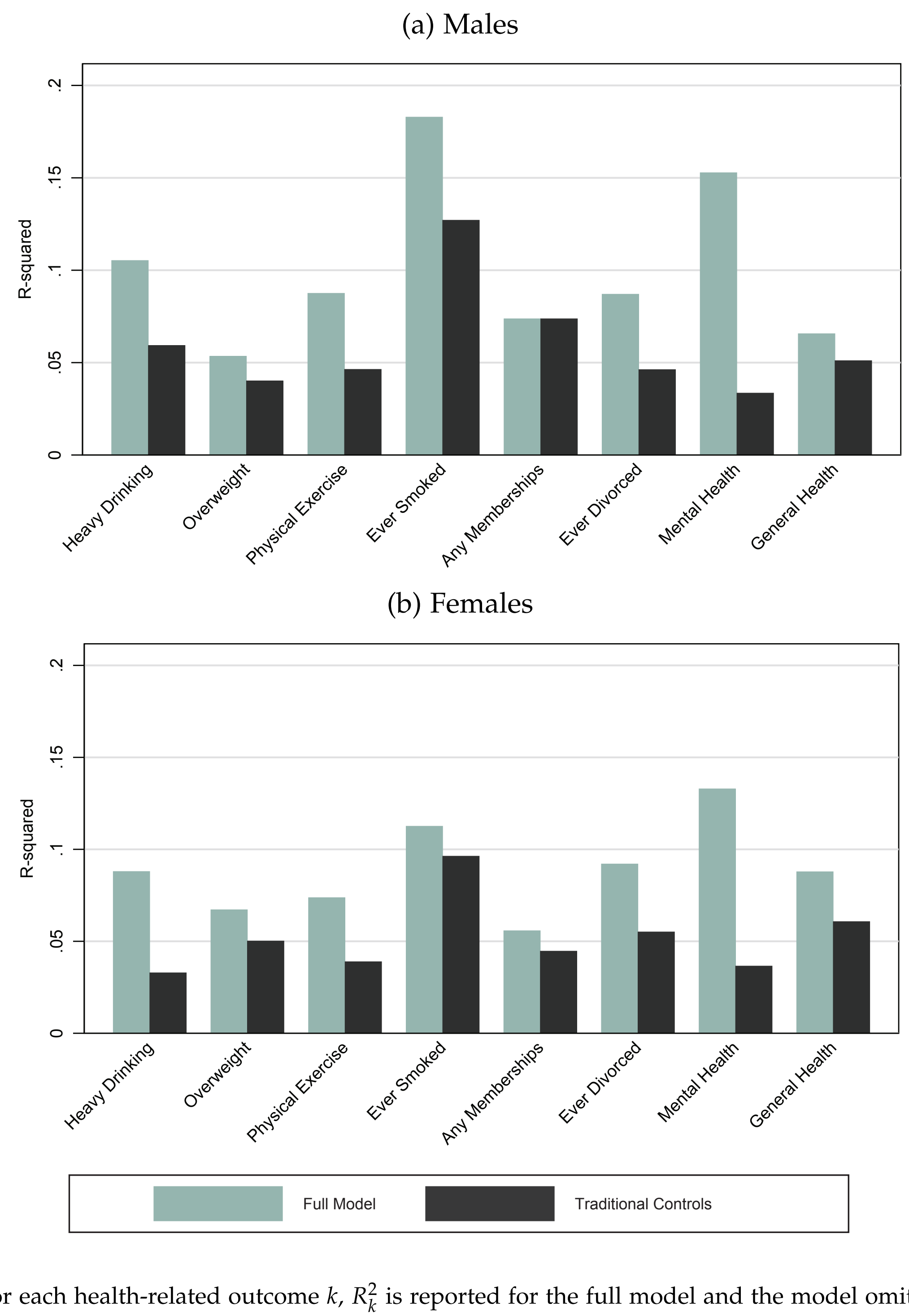
Notes: Each cell in the table shows the regression coefficient representing a conditional association between a skill (or education) and a health-related outcome. The association is conditional on a large set of observable background controls (shown in Table 4) and latent skills. The associations correspond to changes in skills by one standard deviation (for skill coefficients) or to changes in education status from 0 to 1 (for education coefficients). Asterisks denote the stepdown-adjusted statistical significance level within a family (or block) of outcomes of the same type, such as heavy drinking outcomes by year, marked by bold frames. Coefficients are reported with accompanying statistical significance represented by asterisks, where ***,** and * indicate $p < 0.01$, 0.05, and 0.10 respectively. A coefficient with no asterisk refers to $p < 0.15$, while a blank cell refers to a coefficient with p-value above 0.15. Coefficients with p-values above 0.15 are not shown in this summary table, but are available in Tables A-1–A-6 of the Web Appendix, where we present the full set of coefficients, standard errors, and p-values, both adjusted and unadjusted. The results are typeface coded so that bolded coefficients refer to associations that are considered in the literature to be beneficial for longevity (such as a decrease in heavy drinking or an increase in physical activity), and italicized coefficients refer to adverse associations. Calculations are based on the Terman data.

Table 7: Summary of Lifetime and Midlife Outcomes

	Conscientiousness	Openness	Extraversion	Agreeableness	Neuroticism	IQ	Education
A. Lifetime Outcomes, Males							
1940–1960 Ever Drank Heavily	**-.055** *		*.061* *				**-.109** *
1991 Ever Smoked	**-.107** *						
1940–1960 Any Organization							**.084** ***
Ever Divorced	**-.055** *						**-.137** ***
Ever Poor/Fair MH	**-.071** **	*.085* ***			*.134* ***		
Never Poor/Fair GH							
B. Midlife Outcomes, Males[(a)]							
Drank Heavily	**-.072** **						
# of Organizations						**.327** **	**1.501** ***
Mental Difficulty	**-.080** **	*.091* ***	**-.101** ***		*.120* ***		
General Health					*-.211* ***		
C. Lifetime Outcomes, Females							
1940–1960 Ever Drank Heavily		**-.073** *					
1991 Ever Smoked							
1940–1960 Any Organization							
Ever Divorced							**-.111** *
Ever Poor/Fair MH					*.152* ***		
Never Poor/Fair GH					*-.044* *		**.116** **
D. Midlife Outcomes, Females[(a)]							
Drank Heavily							
# of Organizations						*-.352* **	**1.213** ***
Mental Difficulty					*.123* ***		
General Health					*-.241* ***		

Notes: Each cell in the table shows the regression coefficient representing a conditional association between a skill (or education) and a health-related outcome. The association is conditional on a large set of observable background controls (shown in Table 4) and latent skills. The associations correspond to changes in skills by one standard deviation (for skill coefficients) or to changes in education status from 0 to 1 (for education coefficients). Asterisks denote the stepdown-adjusted statistical significance level within a family (or block) of outcomes of the same type, such as heavy drinking outcomes by year, marked by bold frames. Coefficients are reported with accompanying statistical significance represented by asterisks, where ***,** and * indicate $p < 0.01$, 0.05, and 0.10 respectively. A coefficient with no asterisk refers to $p < 0.15$, while a blank cell refers to a coefficient with p-value above 0.15. Coefficients with p-values above 0.15 are not shown in this summary table, but are available in Tables A-7 and A-8 of the Web Appendix, where we present the full set of coefficients, standard errors, and p-values, both adjusted and unadjusted. The results are typeface coded so that bolded coefficients refer to associations that are considered in the literature to be beneficial for longevity (such as a decrease in heavy drinking or an increase in physical activity), and italicized coefficients refer to adverse associations. Calculations are based on the Terman data. [(a)]"Midlife" corresponds to age around 50 based on measurements in 1960.

Figure 1: Coefficient of Determination (R^2) Comparison



Notes: For each health-related outcome k, R_k^2 is reported for the full model and the model omitting latent socioemotional skills. R_k^2 is defined as $1 - V_{\epsilon^k}/V_{H^k}$, where V_{ϵ^k} is the residual variance identified from the factor model in the outcome equation k. V_{H^k} is the outcome variance. Calculations are based on the Terman data.