Understanding the Education Polygenic Score and Its Interactions with SES in Determining Health in Young Adulthood*

Atticus Bolyard† Peter A. Savelyev‡

College of William & Mary College of William & Mary and IZA

February 28, 2020

*A version of this paper was presented to the 28th European Workshop on Econometrics and Health Economics in Leuven, Belgium, and to the 89th Annual Southern Economic Association Meeting in Fort Lauderdale, Florida, USA. The authors are grateful to participants of the workshop for useful suggestions and stimulating discussion. We thank Govert Bijwaard, Gabriella Conti, and Matthew Harris for productive comments. Maxwell Sacher provided excellent proofreading. Bolyard and Savelyev benefited from the research support from The College of William & Mary. The views expressed in this paper are those of the authors and do not necessarily reflect the views of the funders. This research uses data from Add Health, a program project directed by Kathleen Mullan Harris and designed by J. Richard Udry, Peter S. Bearman, and Kathleen Mullan Harris at the University of North Carolina at Chapel Hill, and funded by grant P01 HD31921 from Eunice Kennedy Shriver National Institute of Child Health and Human Development (NICHD), with cooperative funding from 23 other federal agencies and foundations. Add Health GWAS data were funded by NICHD Grants R01 HD073342 (Harris) and R01 HD060726 (Harris, Boardman, and McQueen). Investigators thank the staff and participants of the Add Health Study for their important contributions.

†Address: 300 James Blair Drive, Tyler Hall Room 317, Williamsburg, VA 23185. E-mail: jabolyard@email.wm.edu.

‡Corresponding Author. Address: 300 James Blair Drive, Tyler Hall Room 317, Williamsburg, VA 23185. E-mail: pasavelyev@wm.edu.

Abstract

We investigate the education polygenic score (PGS), an index based on genetic data that predicts years of formal education. Based on the National Longitudinal Study of Adolescent to Adult Health, we find that an education PGS has a beneficial effect on multiple health-related outcomes in young adulthood both indirectly though education and directly. In addition, we find that socioeconomic status (SES) moderates the relationship between the PGS and health outcomes so that individuals with particularly low SES typically do not experience health benefits from the education PGS. We decompose the effect of the PGS with respect to education, early health, and cognitive and noncognitive skills and find that the beneficial effects of PGS work through formal education, early health, and early skills, among other channels. Finally, after controlling for education PGS and unobserved heterogeneity on top of more traditional controls, we still find that education is associated with better health outcomes, which adds evidence to the debate about the causal link between education and health.

Key words: Education PGS, SES, gene-environment interaction, health, health-related outcomes

JEL codes: I12, I14, I24, J24

1 Introduction

There is a quickly growing availability of genetic data, creating new research opportunities for economists. We study a measure of genetic endowment that predicts education—an education polygenic score (PGS). PGSs are popular due to their strong predictive power and more robust results across populations than for specific genes (Benjamin et al., 2011). Yet we still know little about how PGSs interact with the environment, the predictive power of PGSs outside the outcomes they are designed to predict, and the mechanisms through which they affect outcomes. An education PGS, which measures an individual's genetic predisposition for completing formal schooling, has particularly important socioeconomic implications. This paper studies the effect of an education PGS on health-related outcomes, its interaction with socioeconomic status (SES) in childhood, and the mechanisms behind these effects.

We have four main contributions. First, we show that an education PGS predicts health and a variety of health-related behaviors not only through formal education but also directly.[1] One practical implication from this finding is that an education PGS does not make a valid instrument for the effect of education on health contrary to a growing literature using this instrument. Our result holds for a variety of health outcomes after controlling for cognitive and noncognitive skills and unobserved heterogeneity. Second, we show that the education PGS exhibits differential effects depending on childhood SES: for a number of health outcomes we observe beneficial effects of the education PGS on health for high-SES subjects only.[2] Third, our mediation model reveals the mecha-

[1]Even the indirect effect of the PGS on health through education is not guaranteed ex ante given the fact that the causal status of the relationship between education and health-related outcomes is controversial, as we discuss in Section 2.

[2]We focus on measures of SES that imply major socioeconomic hardship in childhood to examine the effect of an SES "bottleneck." Our SES dummy takes the value of 1 if the respondent's family experienced at least one of three proxies for low SES in childhood: living in an unsafe neighborhood, receiving government assistance, or having trouble paying bills.

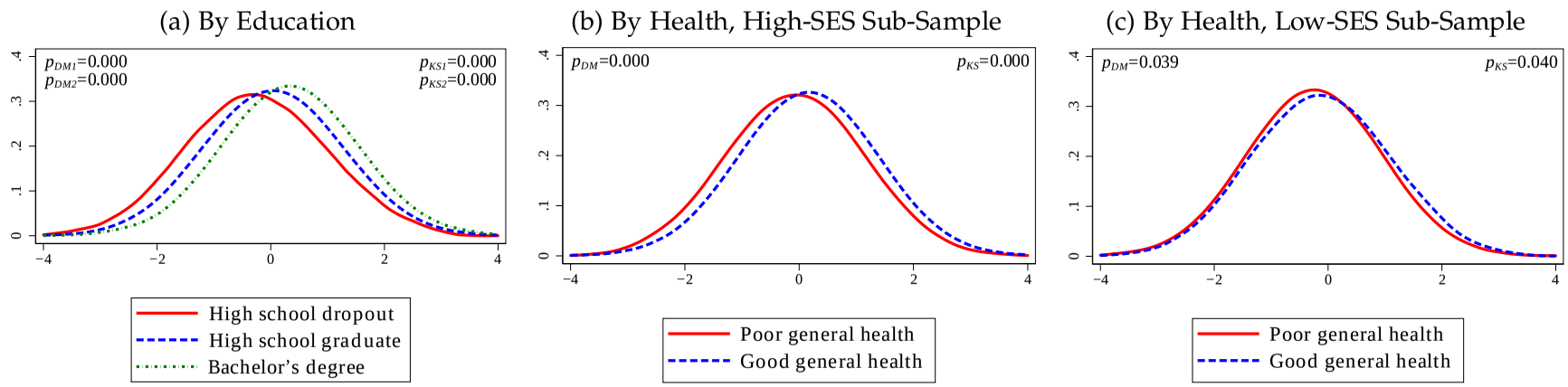
nisms through which the education PGS works, which allows us to better understand the score itself. Finally, we contribute to the debate on the causal effect of education on health by ruling out important genetic confounders.

We demonstrate our results non-parametrically in Figure 1. Panel (a) shows that the education PGS has a statistically significant difference by education level both distributively and on average. So, the education PGS indeed predicts formal education. Panel (b) shows that for the high-SES sub-sample, a higher polygenic score is associated with better health in young adulthood. Panel (c) shows that the same association is weaker for the low-SES sub-sample. Later, once we control for both observed and unobserved heterogeneity in our main model, we still find a strong association between the education PGS and general health for the high SES group, but do not find even a weak association for the low-SES group.

We are not the first to study the effects of an education PGS on outcomes other than education. Barth et al. (2018) find that an education PGS also explains large amounts of variation in stock market returns and wealth inequality. They offer suggestive evidence that an education PGS may capture aspects of individual information-processing that affects decisions, such as conceptions of probability and risk aversion. This evidence is in line with our results, as better planning and risk aversion are complementary with health. Additionally, Belsky et al. (2016) find that the score is related to several socioeconomic outcomes, as well as various skill dimensions, but do not find associations with health. Our paper broadens the set of known outcomes that are affected by the education PGS by showing its effects on multiple health-related outcomes in young adulthood. Economists have also examined the impact of education as a moderator of genetic effects itself, namely the effect of a mental health PGS on depression (Amin et al., 2019).

Our results also contribute to better understanding of Mendelian Randomization (MR), a technique of using genetic variables as instruments for the traits they predict. Already, studies in several fields, including economics, have made use of the score as

Figure 1: Kernel Densities of Education PGS by Education and Health in Young Adulthood



Notes: Calculations are based on the AddHealth Data. Education PGS is standardized for the full estimation sample. Epanechnikov Kernel Densities are used with bandwidth 0.7. "Poor health" is defined as responding "poor," "fair," or "good" when asked to evaluate ones own health. "Good health" is defined as "very good" or "excellent" reported health. We test the equality for differences in means (DM) (see p-values p_{DM}) and the equality of distributions using the Kolmogorov-Smirnov (KS) test (see p-values p_{KS}). In Panel (a) DM and KS tests 1 and 2 compare Bachelor's degree holders with high school graduates and high school dropouts respectively. In Panels (b) and (c) DM and KS tests compare those with poor health to those with good health. Low SES is defined as any of the following being true: family receiving government assistance, having trouble paying bills, or living in an unsafe neighborhood.

an instrument for estimating the effect of education on health outcomes (Böckerman et al., 2017; Davies et al., 2019; Tillmann et al., 2017). By showing that the education PGS has direct effects on an array of health outcomes, including general health, we provide evidence that the exclusion restriction is violated in these studies, undermining their explanatory potential (Munafò et al., 2019; von Hinke et al., 2016). Our specific findings about Education PGS as an invalid instrumental variable for health outcomes are in line with the general criticism of the MR approach by DiPrete et al. (2018), who offer a method of estimating the upper and lower bounds of the causal effect that works even in the case when the instrument is invalid.

We use data from The National Longitudinal Study of Adolescent to Adult Health (Add Health), which follows a cohort of individuals who were in either middle or high school in 1996, and who are now young adults. We study a variety of health outcomes related to general and mental health, substance use, exercise, and body weight. We only study whites due to limited genetic data availability for non-whites.

We start our empirical analysis by estimating a reduced form model similar to other models of gene-environment interaction in the literature by including an interaction term between a PGS and the environment (e.g., Bierut et al., 2018). In this model we find associations between the education PGS and health-related outcomes. Another novel result is a substantial interaction between the education PGS and childhood SES. This interaction implies that the education PGS tends to predict better health only for those with high SES. Thus, a reduced-form parametric model shows results that are similar to nonparametic results from Panels (b) and (c) of Figure 1. This result is in line with the environmental bottleneck hypothesis, which suggests that adverse environments can limit the benefits of productive genes (Fletcher, 2019). Fletcher finds that poor parental health can limit the effects of a genetic predisposition towards educational attainment on actual education. Schmitz and Conley (2017) find that reductions in educational attainment as a result of Vietnam-era conscription are larger for individuals with lower edu-

cation polygenic scores, providing evidence that a combination of experiencing severe environmental conditions and having an unfavorable genetic endowment is particularly harmful.

However, we take this framework a step further in our main model by accounting for mechanisms behind these effects and for endogeneity. Results of the main model confirm the bottleneck role of childhood SES for health development detected in the reduced-form model and measure the role of early skills and early health as the mechanisms behind this effect. In order to model the mechanisms we estimate a set of recursive equations that model skills, early health, education, and health outcomes. We explicitly control for multidimensional skills and for unobserved heterogeneity.

It is a common concern in the literature that effects of genetic scores are confounded by family background, since parents and children share some of their genetic endowments. For instance, some aspects of intelligence, which we show is captured in the education PGS, are heritable. Belsky et al. (2016) find that high-SES families tend to have children with higher education polygenic scores. However, Lee et al. (2018) find that the education PGS still predicts educational attainment among siblings after controlling for family fixed effects. We minimize these concerns by explicitly controlling for a large set of background variables, as well as census tract-level crime, education, unemployment, and poverty rate. We also perform an alternative estimation using family fixed effects based on 230 families who have at least two children surveyed by Add Health. Similar to Amin et al. (2019), who study a mental health PGS based on the same Add Health dataset, we find the family fixed effect approach to be greatly underpowered.

Our model also contributes to a separate body of literature on the relationship between education and health, the causal status of which is still debated (see Section 2). One possible confounding factor in education-health studies is genetic endowment, which may be strongly predictive of both education and health (Boardman et al., 2015). We find that even after controlling for the education PGS, which we show to be predictive

of both education and health, education still exhibits large and statistically significant associations with a variety of health outcomes. Accounting for a major potential confounder in addition to other important controls strengthens the evidence that education has a causal effect on health.

2 Related Literature

In this section we continue discussing our paper in light of the literature.

Some results in the literature suggest that there might be a relationship between education polygenic scores and health. Marioni et al. (2016) suggests that an education PGS may contribute to health, but the authors' evidence regarding this specific question is indirect. The authors use data on European cohorts and show that a child's education PGS is predictive of parental longevity. They suggest that some genetic mechanism captured in a child's education PGS might be relevant to determining parental health, since parents and children have common genetic traits.

In a related paper, Barcellos et al. (2018), who focus on the role of education in reduction of genetic risk of obesity, find a weak association between the education PGS and blood pressure, as well as an association between the education PGS and a weighted average of blood pressure, body size, and lung function. Unlike this paper, the authors do not study general health and health behaviors, do not interact PGS with SES, and do not decompose the effects of PGS with respect to their mediators.

Understanding the relative contributions of genetic endowment, environment, and gene-environment interactions to health outcomes is crucial to understanding health capital formation (Heckman, 2007), which is a fundamental economic question (e.g., Galama and Van Kippersluis, 2018; Grossman, 2006). Among possible environmental factors, SES can be expected to be a strong and policy-relevant environmental influence (Strauss and Thomas, 2007). Additionally, previous work has shown that SES moderates

the effects of a number of polygenic scores. Bierut et al. (2018) show that SES moderates the effect of a smoking PGS on actual smoking outcomes. Papageorge and Thom (2018) find that an education PGS is associated with higher education gains when children have high SES. However, to the best of our knowledge, we are the first to study the interaction between an education PGS and childhood SES in predicting health outcomes.

Avinun (2019) finds a mediating role of an individual's SES in the positive association between education PGS and lower depression. Our paper has a different focus, as we study an interaction of PGS with childhood SES (which is parental SES at the time of child development) as a measure of a child's developmental bottleneck rather than mediation though a person's own SES later in life.[3] Also, we study not only depression but seven other health outcomes as well as education.

Cognitive and noncognitive skills are important determinants of both health-related outcomes and education (e.g., Cunha and Heckman, 2007). Cognitive skills have been shown to predict better health (e.g., Auld and Sidhu, 2005; Bijwaard et al., 2015). Also, economists have begun to pay increasing attention to noncognitive skills, and have shown that they are important determinants of health-related outcomes (e.g., Conti et al., 2010; Savelyev and Tan, 2019). We examine whether the effects of the PGS on health works through skills and find evidence of such an effect.

We also contribute to the debate about the effect of education on health. Apart from regressions conditional on observable controls and propensity score methods, there are three major methods used to identify the causal effect of education on health-related outcomes: (1) the use of randomized or natural experiments as sources of exogenous variation; (2) the use of family fixed effects or twin fixed effects; (3) the explicit modeling of unobserved heterogeneity.[4] All these methods have their advantages and disadvan-

[3]We partly address individual SES mediation as well, as individual education can be viewed as a measure of individual SES.

[4]This approach is also referred to as "structural," as it assumes specific structure linking observed and latent variables.

tages, as discussed below. Literatures (1) and (2) lead to results that contradict each other. Literature (3) tends to find effects of education on health, but the concern with this literature is whether all confounders are fully controlled. This paper diminishes concerns about results from literature (3) by adding a powerful genetic confounder in addition to detailed observable controls, cognitive and noncognitive skills, and accounted unobserved heterogeneity, and still finds strong associations between education and health. Below we briefly explain results of literatures (1–3).

In literature (1), the use of randomized controlled experiments in education is usually limited to early childhood education (Conti et al., 2016). Natural experiments have a well-defined source of variation, but they only identify the local treatment effect and may suffer from lack of validity, lack of monotonicity, and weakness of instruments (e.g., Heckman and Vytlacil, 2007). Literature (1) mostly relies on changes in compulsory schooling laws as a source of exogenous variation, though rare exceptions exist, like the use of military draft avoidance (Buckles et al., 2016). Results of these papers differ greatly. For instance, some (e.g., Lleras-Muney, 2005; van Kippersluis et al., 2011) find a strong effect of education on health-related outcomes, while others (e.g., Albouy and Lequien, 2009; Clark and Royer, 2013; Mazumder, 2008; Meghir et al., 2018) find none. Likely reasons for these differences include weakness of compulsory schooling laws as an instrument for a number of countries including the US, confounding influences of other reforms, and differences in effects by population, cohort, and sex (Galama et al., 2018).

Literature (2) relies on fixing a large number of unobserved confounders that are shared by twins or siblings. However, estimates based on these methods are highly sensitive to measurement error in education (e.g., Ashenfelter and Krueger, 1994) and could be confounded by unobserved health differences among siblings or twins in their early life. Also, external validity of results based on these methods might be challenging. Just as for literature (1), there are contradictory results for literature (2). Some papers

find substantial effects (e.g., Lundborg et al., 2016; Savelyev et al., 2019; van den Berg et al., 2015), while some other find little to no effects (e.g., Amin et al., 2015; Behrman et al., 2011; Madsen et al., 2010). Differences in results could be partly related to different model specifications and partly due to differences by population, cohort, and sex.

Literature (3) explicitly models relationships between observed and unobserved confounders, education, and health-related outcomes. These methods tend to better preserve statistical power than methods (1) and (2), and, unlike (1), do not limit results to local average treatment effects. The biggest concern with this literature is its ability to account for all important confounders.

Important confounders that are explicitly accounted for in literature (3), usually though factor modeling to account for measurement error, include major human capabilities: health, cognitive skills, and noncognitive skills (Bijwaard et al., 2015; Conti and Heckman, 2010; Savelyev and Tan, 2019). Savelyev (2019) and Hong, Savelyev, and Tan (2019) also account for latent unobserved heterogeneity in addition to latent human capabilities. The innovation of this paper is that in addition to controls that have been used in this literature, we account for the education PGS, a strong genetic confounder, and still find strong conditional associations between education and health-related outcomes.

3 Data

We use the National Longitudinal Study of Adolescent to Adult Health (Add Health). This panel dataset follows roughly 20,000 individuals and contains detailed information on their family background, health outcomes, skills, and education. The respondents were first surveyed in 1995–1996, when they were in grades 7–12, and were followed into young adulthood. The most recent data that are used in this paper, Wave IV, were collected when participants were 24–32 years old.

Add Health participants were drawn from a sample of selected high schools. High schools were randomly chosen by stratifying schools within the Quality Education Database according to several demographic factors, and weighting the probability that they would be selected according to their enrollment. One middle school feeder school was randomly selected for each selected high school, weighted proportionally to its size. Respondents were chosen randomly after being stratified according to grade, sex, and school. The sample is considered nationally representative (Harris, 2013).

Our sample size is constrained by the availability of genetic data. About 9,000 Add Health study participants took part in genotyping, 5,728 of whom are white. We perform this analysis only for individuals who self-identify as white due to data limitations described below.

Education and Health Outcomes We study three education thresholds: high school dropout, high school graduate, and college graduate. Previous work on health returns to education has shown that having a Bachelor's degree is particularly predictive of health (Buckles et al., 2016; Savelyev, 2019). Distinguishing between high school dropouts and high school graduates allows us to quantify health effects of the PGS for individuals through an additional education margin. There is extensive evidence that high school graduation has substantial health benefits (Freudenberg and Ruglis, 2007).

All of the health outcomes in this study are measured in Wave IV. Self-reported health has been shown to be predictive of mortality, and it is an essential measure of overall health (Idler and Benyamini, 1997). We also study three binary variables related to substance use: heavy marijuana use, smoking tobacco, and heavy drinking. Finally, we analyze two lifestyle variables, "obesity" and "no exercise," and two mental health variables, "having suicidal thoughts" and "depression." See Table 1 for variable definitions and descriptive statistics.

Table 1: Education PGS, Education, and Health-Related Outcomes in Young Adulthood

	Full sample			Low SES			High SES		
	Average	Standard deviation	Sample size	Average	Standard deviation	Sample size	Average	Standard deviation	Sample size
Education									
High school dropout	0.044	0.205	3,813	0.081	0.272	1,215	0.027	0.161	2,598
High school degree	0.638	0.481	3,813	0.747	0.435	1,215	0.587	0.492	2,598
Bachelor's degree or above	0.318	0.466	3,813	0.173	0.378	1,215	0.386	0.487	2,598
Polygenic Score									
Education PGS	0.000	1.000	3,813	-0.176	0.979	1,215	0.082	0.999	2,598
Early Health									
General health[(a)]	3.917	0.864	3,813	3.784	0.902	1,215	3.980	0.838	2,598
Health-Related Outcomes									
General health[(a)]	3.732	0.889	3,813	3.556	0.918	1,215	3.815	0.863	2,598
Obesity[(b)]	0.344	0.475	3,768	0.398	0.490	1,197	0.320	0.466	2,571
No exercise[(c)]	0.136	0.343	3,654	0.159	0.366	1,164	0.126	0.332	2,490
Heavy drinking[(d)]	0.259	0.438	3,795	0.285	0.452	1,205	0.247	0.431	2,590
Smoking[(e)]	0.395	0.489	3,800	0.215	0.411	1,211	0.159	0.366	2,589
Marijuana use[(f)]	0.077	0.266	3,810	0.096	0.295	1,214	0.067	0.251	2,596
Depression[(g)]	0.192	0.394	3,813	0.215	0.411	1,215	0.182	0.386	2,598
Suicidal thoughts[(h)]	0.073	0.261	3,804	0.096	0.295	1,210	0.063	0.243	2,594

Notes: Calculations are based on the AddHealth data. [(a)]Raw variable is shown, but it was standardized prior to analysis. General Health is measured on a scale from 1 (poor) to 5 (excellent). [(b)]Obesity is defined according to the standard Body Mass Index-based definition, which is used by the National Institute of Health (BMI $\geq$ 30). [(c)]No exercise equals one if the respondent reports none of the following: playing sports, exercising outside, walking for exercise, or engaging in other physical activity during the past week. It is zero otherwise. [(d)]Heavy drinking is defined as consuming 4 or more drinks per drinking occasion as a female, and 5 or more drinks per occasion as a male, which follows the Centers for Disease Control and Prevention's definition of binge drinking (National Institute on Alcohol Abuse and Alcoholism, 2004). [(e)]Smoking is defined as smoking cigarettes at least once within the past 30 days. [(f)]Marijuana use is defined as smoking marijuana more than once a week, on average, during the last year. [(g)]Depression equals one if the respondent had ever been told by a health care provider that they have depression. [(h)]The suicidal thoughts variable equals one if the respondent reported seriously considering suicide within the last year.

Polygenic Scores Although the effect of a genetic endowment is contingent on environmental factors, such as family characteristics and educational institutions, polygenic scores are still predictive across environments, however some heterogeneity in effect sizes exists. We use a PGS constructed by Lee et al. (2018) using information from 1,271 individual genetic characteristics that are predictive of education[5] and based on a sample of 1.1 million individuals of European descent. The PGS predicts about 11% of variation in years of education completed among a validation sample that includes Add Health data (Lee et al., 2018). Given that the formula for this education PGS was determined with no use of the Add Health Dataset, Panel (a) of Figure 1 is further proof of external validity and high predictiveness of the education PGS.

A major methodological concern is that a PGS may exhibit bias, or provide decreased statistical power, if population stratification is not controlled for, since the functions of genes vary by anthropological background (Price et al., 2006). Therefore, our study is based only on those who self-report being "white," since the education PGS is constructed from a sample of individuals with European ancestry. Additionally, we control for principal components of genetic ancestry. This is a standard method of controlling for population stratification, which is recommended for Add Health users. Specifically, Add Health suggests using the first five principle components, which we follow (Harris and Braudt, 2018).

Polygenic scores are constructed by regressing a large set of individual genetic characteristics on an outcome of interest, and keeping only coefficients with very low p-values in order to control for the familywise error rate. The coefficients are then used to weight the contribution of each single genetic variation to the overall PGS. For each individual, a PGS is constructed by creating a weighted sum all of the relevant genes an

[5]The genetic characteristics are single-nucleotide polymorphisms (SNPs), which are nucleotides at a particular location on the genome that represent variation among humans.

individual possesses (Domingue et al., 2015).[6] We standardize the PGS for the estimation sample prior to model estimation (see Table 1 for descriptive statistics).

SES We create a dummy variable indicating whether an individual reports experiencing at least one of three situations that are indicative of a particularly low family SES: whether they lived in an unsafe neighborhood, whether their family received government assistance (such as welfare), and whether their family had difficulty paying bills. About 32% of the sample has low SES according to this definition, which is a large population (see Table 2). The qualitative measures we use are intended to capture hardship in childhood. This is a similar approach taken by other authors in this literature, including Papageorge and Thom (2018) and Bierut et al. (2018). We use a dummy variable, rather than an index, due to concerns about the linearity of effects. We show in Section 5 that this measure of SES has advantages over family income.

Cognitive and Noncognitive Skills Although there are multiple alternative taxonomies for representing noncognitive skills, one well-established and comprehensive representation of noncognitive skills is the Big Five Personality taxonomy. Most noncognitive skills map into the Big Five in some manner (Borghans et al., 2008). The Big Five skills are Openness, Conscientiousness, Extraversion, Agreeableness, and Emotional Stability. Openness is a propensity to be open to new experiences and ideas; Conscientiousness is a propensity to follow rules and plan the future; Extraversion is a propensity to be active and social; Agreeableness is a propensity to behave friendly towards others; and Emotional Stability is a propensity to control one's emotions. Since skills are malleable over the lifecycle (e.g., Borghans et al., 2008; Fletcher and Schurer, 2017), we model skills measured during the first wave of the Add Health study to avoid capturing reverse causality.

[6]The linkage disequilibrium regression method is used to avoid double counting correlated genetic effects.

Table 2: SES Measures and Background Controls

	Full sample		Low SES		High SES	
	Average	Standard deviation	Average	Standard deviation	Average	Standard deviation
Low SES Measures						
Unsafe neighborhood	0.064	0.245	0.202	0.401	0.000	0.000
Received assistance	0.210	0.407	0.666	0.472	0.000	0.000
Trouble paying bills	0.121	0.327	0.381	0.486	0.000	0.000
Low SES dummy	0.319	0.466	1.000	0.000	0.000	0.000
Background Controls						
College-educated parent	0.521	0.500	0.434	0.496	0.552	0.497
Cigarettes smoked in house	0.471	0.499	0.588	0.492	0.417	0.493
Family income (log)	3.705	0.762	3.284	0.812	3.896	0.654
Only child	0.045	0.208	0.040	0.195	0.048	0.214
First born	0.319	0.466	0.283	0.451	0.336	0.472
Second born	0.308	0.462	0.281	0.450	0.321	0.467
Third born	0.117	0.322	0.125	0.331	0.114	0.317
Number of siblings	2.548	1.942	3.000	2.277	2.337	1.724
Low birth weight	0.086	0.280	0.095	0.293	0.081	0.273
Rural	0.365	0.482	0.373	0.484	0.361	0.480
Suburban	0.392	0.488	0.332	0.471	0.421	0.494
Urban	0.223	0.417	0.260	0.439	0.206	0.405
Parents married	0.796	0.403	0.642	0.480	0.867	0.340
Meals with parents[(a)]	4.987	2.307	4.775	2.451	5.086	2.230
South	0.362	0.481	0.404	0.491	0.342	0.475
Northeast	0.154	0.361	0.157	0.364	0.153	0.360
West	0.145	0.352	0.120	0.325	0.156	0.363
Midwest	0.339	0.474	0.319	0.466	0.349	0.477
Male	0.473	0.499	0.448	0.497	0.484	0.500
Age Wave I: 10-12	0.098	0.297	0.097	0.296	0.098	0.297
Age Wave I: 13-14	0.325	0.469	0.336	0.472	0.320	0.467
Age Wave I: 15-16	0.384	0.486	0.395	0.489	0.379	0.485
Age Wave I: 17-19	0.189	0.391	0.166	0.372	0.199	0.399
Hispanic	0.061	0.240	0.099	0.298	0.045	0.207
Genetic ancestry PC[(b)]	Yes		Yes		Yes	
Census Regional Variables						
Crime rate (per 100,000)	4724	2540	4776	2792	4700	2414
Education rate[(c)]	0.227	0.130	0.198	0.115	0.241	0.134
Unemployment rate	0.064	0.024	0.068	0.025	0.063	0.023
Poverty rate	0.115	0.091	0.144	0.106	0.101	0.079
Estimation Sample Size	3,813		1,215		2,598	

Notes: Calculations are based on the AddHealth data. We control for five principal components describing ethnic origin. [(a)] Number of evening meals. [(b)] Five principle components (PC). [(c)] Percentage of adults with college degree.

We use the factor model described in Section 4. Due to data limitations, we are not able to study early Agreeableness and Openness, but we account for early Conscientiousness, Extraversion, Emotional Stability, and Intelligence using measures obtained when the Add Health cohort was in grades 7–12. We used the same measures as proposed by psychologists using Add Health data (Young and Beaujean, 2011).

We also use three measures of cognitive skills to construct factor scores for Cognition. We use participant scores on the Add Health Picture Vocabulary Test, recent science grades, and recent math grades as measures for intelligence. The Add Health Picture Vocabulary test is a shortened version of the Peabody Picture Vocabulary Test. See Table A-1 in the Appendix for the list of measures of each cognitive and noncognitive skill.

Controls for Family Background, Geographic Information, and Initial Health We control for a range of background variables that could influence education and health. These variables describe parental education, smoking in the household, family income, birth order, family structure, birth weight, degree of urbanization, marital status of parents, cohort, and ancestry. Additionally, we control for several tract-level census variables: the crime rate, education rate, unemployment rate, and poverty rate of the respondent's census-tract. Together, these variables capture a large amount of information about the respondent's family and social environment. Lastly, we control for biological sex (male or female), as reported in Wave I. See Table 2 for descriptive statistics.

4 Methods

We first estimate a reduced form model of the PGS-SES interaction, comparable to the economic literature on gene-environment interactions (e.g., Bierut et al., 2018; Fletcher,

2019; Schmitz and Conley, 2017). The model is specified as follows:

$$H_k^* = \alpha_{1k} + \alpha_{2k} PGS + \alpha_{3k} SES + \alpha_{4k} PGS \cdot SES + \alpha_{5k} \tilde{\boldsymbol{X}} + \zeta_k, \quad k = 1, ..., 8, \tag{1}$$

where H_k^* denotes a latent propensity for health outcome H_k in a logit model, so that $H_k = 1$ if $H_k^* > 0$ and $H_k = 0$ otherwise. For continuous outcomes $H_k^* = H_k$. PGS denotes the education PGS; SES, the binary low SES status; and ζ_k, an idiosyncratic error term. The vector of controls $\tilde{\boldsymbol{X}}$ is limited in this model to principal components, region variables, cohort variables, and sex to make this specification comparable with the literature (Bierut et al., 2018).

In addition to model (1), we also estimate our main model. We start with a factor model for early skills. We follow a paper by psychologists Young and Beaujean (2011) who propose measures of early Conscientiousness (C), Extraversion (E), and Emotional Stability (ES) based on available measures of personality in the first wave of Add Health. We also estimate a factor of early Cognition (G) based on a cognitive test and school achievements.[7] We predict Barlett's factor scores as estimates of skills. Factor scores provide estimates of latent variables and can be used for statistical inference when properly estimated (Anderson and Rubin, 1956).[8]

The factor model is specified as follows:

$$M_j^s = \beta_{0,j}^s + \beta_{1,j}^s \Theta^s + \eta_j^s, \quad j = 1, ..., J^s, \tag{2}$$

where Θ^s is a latent skill, such that $s \in \{C, E, ES, G\}$; M_j^s is the j-th measure of skill s; J^s is the total number of measures of skill s; $\beta_{0,j}^s$ is the intercept, $\beta_{1,j}^s$ is the factor loading,

[7] See Table A-1 of the Appendix for the list of measures of skills. We reanalyzed measures by Young and Beaujean (2011) using our own exploratory and confirmatory factor analysis and confirm that their model is well-specified.

[8] We choose factor scores over latent factors for practical reasons. The choice hardly affects results but greatly decreases the computational intensity of the model, which is computationally intensive even with the use of factor scores.

and η_j^s is the error term. We make the standard assumption that $\eta_j^{s_1} \perp\!\!\!\perp \eta_l^{s_2}$ for any $s_1, s_2 \in \{C, E, ES, G\}$ and any pair of j and l, $j = 1, ..., J^s$; $l = 1, ..., J^{s_2}$, except for the case where $s_1 = s_2$ and $j = l$, and that error terms are independent of latent factors. Each latent variable is normalized to have mean zero. For each type s of a factor, we make a standard normalization: $\beta_{1,1}^s = 1$. In marking our notarizations we make sure that skills have the usual interpretation without a loss of generality.[9] From this factor model we calculate a vector of Bartlett factor scores. Bartlett's method provides unbiased factor score estimates based on maximum likelihood techniques (Yung and Yuan, 2013).

We simultaneously estimate a model of initial skill and health formation, education choices, and a health outcome, using a full maximum likelihood estimation. We take advantage of the sequential nature of early skill and initial health formation, schooling decisions, and health outcomes in young adulthood. Genes are determined at conception, family SES in childhood is determined by parental choices and background, early skills and initial health are formed before schooling decisions are made, and schooling occurs before adult health outcomes are realized. It is plausible that later outcomes do not affect early outcomes conditional on background controls and unobserved heterogeneity. These exclusion restrictions lead to recursive equations. Simultaneously estimating a system of equations provides both gains in efficiency and allows us to control for unobserved heterogeneity, μ. We control for unobserved heterogeneity using the semi-parametric heterogenity model (Heckman and Singer, 1984).[10] Mroz (1999) shows that such control for unobserved heterogeneity is effective for solving the endogeneity problem in a recursive nonlinear system of equations. We interpret this control for unobserved heterogeneity as capturing important personal, family, or genetic characteristics outside of what is explicitly controlled for.

[9] For instance, more emotionally-stable people have a higher (not lower) score for "Emotional Stability."

[10] The model is also known as latent class analysis (Aitken and Rubin, 1985), discrete factor approximation (Mroz, 1999), discrete factor random-effects model (Gilleskie, 2014), and finite mixture modeling of unobserved heterogeneity (Cameron and Trivedi, 2005).

The sub-model for outcomes is the following:

$$H_0 = a_0 + a_1 PGS + a_2 SES + a_3 PGS \cdot SES + a_4 \boldsymbol{X} + \mu_{H_0} + \epsilon_{H_0} \tag{3}$$

$$\Theta^s = b_0^s + b_1^s PGS + b_2^s SES + b_3^s PGS \cdot SES + b_4^s \boldsymbol{X} + \mu_{\Theta^s} + \epsilon_{\Theta^s}, \quad s \in \{C, E, ES, G\} \tag{4}$$

$$\begin{aligned} D_l^* = {} & c_{0l} + c_{1l} PGS + c_{2l} SES + c_{3l} PGS \cdot SES + c_{4l} \boldsymbol{X} + \boldsymbol{c}_{5l} \boldsymbol{\Theta} + c_{6l} H_0 \\ & + c_{7l} H_0 \cdot SES + \mu_{D_l} + \epsilon_{D_l}, \quad l = 1, 3 \end{aligned} \tag{5}$$

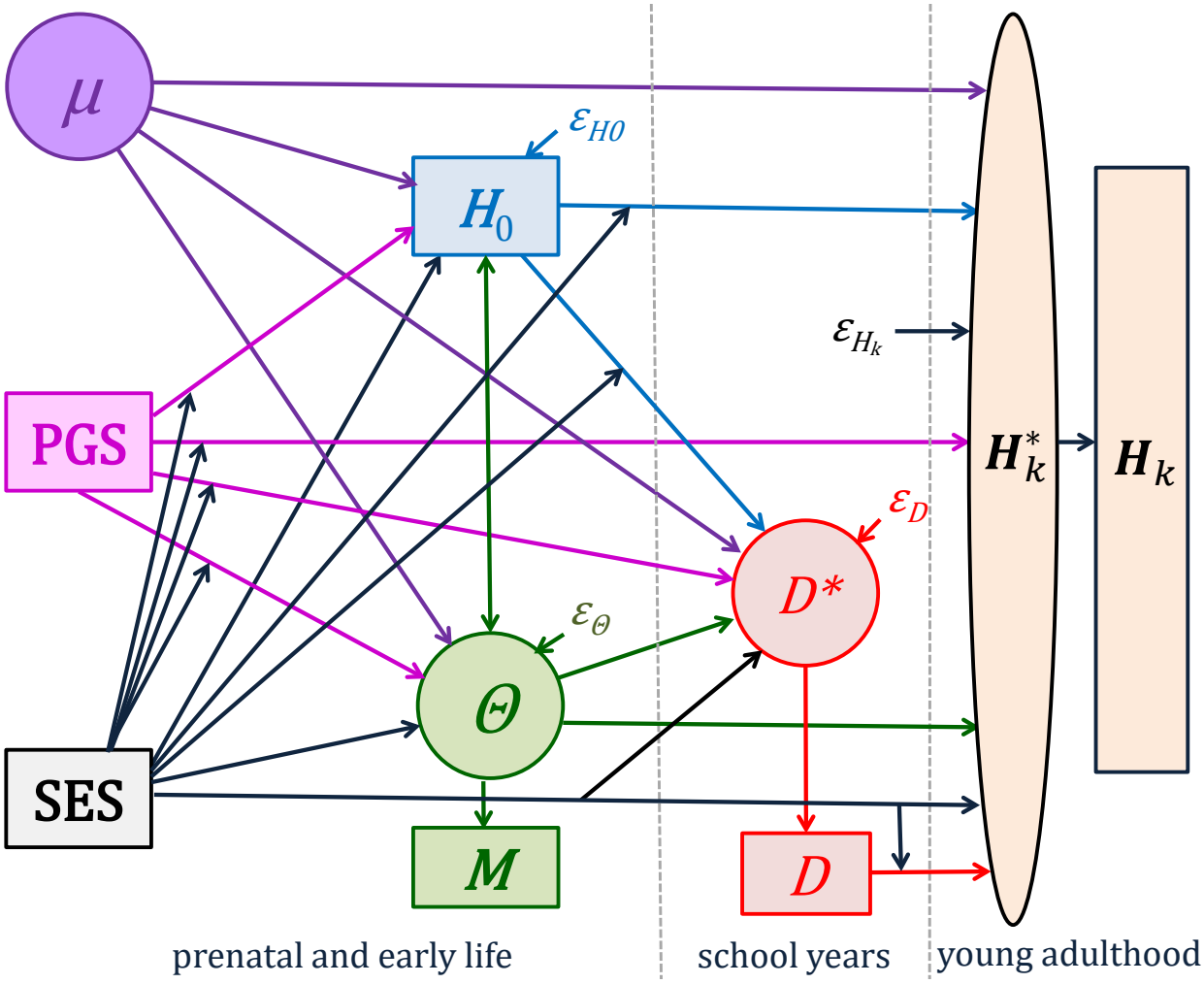
$$\begin{aligned} H_k^* = {} & d_{0k} + d_{1k} PGS + d_{2k} SES + d_{3k} PGS \cdot SES + d_{4k} \boldsymbol{X} + \boldsymbol{d}_{5k} \boldsymbol{\Theta} + d_{6k} H_0 \\ & + d_{7k} H_0 \cdot SES + d_{8k} D_1 + d_{9k} D_1 \cdot SES + d_{10k} D_3 + d_{11k} D_3 \cdot SES \\ & + \mu_{H_k} + \epsilon_{H_k}, \quad k = 1, ..., 8, \end{aligned} \tag{6}$$

where we use the same notation as in model (1), as well as: H_0, early health; $\boldsymbol{\Theta}$, the vector of skills; D_l^*, the latent propensity to dropout from high school (for $l = 1$) or to complete Bachelor's degree or above (for $l = 3$). We do not need to have a separate model for high school graduation (for $l = 2$), as probabilities of these three levels of education sum up to one. Also, high school graduation is the omitted binary variable in Equation (6). We make the standard assumption that unobserved heterogeneity is independent of controls (van den Berg, 2001). Unlike vector $\tilde{\boldsymbol{X}}$ in Equation (1) the vector $\boldsymbol{X}$ represents the complete set of available background controls (see Table 2). Due to computational limitations, we estimate the model separately for each health related outcome H_k. Symbols ϵ_{H0}, ϵ_{Θ}, ϵ_{D_l}, and ϵ_{H_k} represent idiosyncratic error terms.

We present a visualization of our main model (2–6) in Figure 2, which shows causal links between variables from early life, school years, and mid-life, relationships between observed and latent variables, and interaction effects.

This model consists of standard components, identification of which is well established. The factor model (2) is standard and has at least three measures of each factor and all necessary normalizations, which leads to its identification (Anderson and Rubin, 1956). Equations (3–6) have a recursive structure, which contributes to identifica-

Figure 2: Diagram of the Main Model Conditional on Background Variables $\boldsymbol{X}$



Notes: The diagram represents model (2–6) conditional on background variables $\boldsymbol{X}$. Rounds and ellipses represent latent variables. Rectangles represent observables. Arrows ending in a rectangle or round denote causal links. Arrows ending in the middle of another arrow denote interactions. Notation of variables matches the notation in equations (2–6): μ is unobserved heterogeneity; H_0 and $\boldsymbol{\Theta}$ represent early general health and skills. $\boldsymbol{M}$ is a vector of skill measures, which include measures of cognition and personality; D^* is a latent propensity for education; D is chosen highest education level; H_k^* is a latent propensity to have health-related outcome of type k; H_k is health outcome of type k; ϵ_{H0}, ϵ_Θ, ϵ_D, ϵ_{H_k} represent idiosyncratic error terms.

tion conditional on idiosyncratic error terms being uncorrelated among recursive parts (Maddala, 1983). Unobserved heterogeneity terms $\boldsymbol{\mu}$ helps to satisfy this condition by controlling for common correlation across outcomes conditional on observables.

Our main model allows us to follow a standard decomposition technique and decompose the conditional expectation of H_k with respect to mediators of education PGS. In particular, by taking a derivative of the conditional expectation of H_k with respect to the education PGS and applying the chain rule, we can get two parts of the effect of the education PGS on health outcome k: an indirect effect (IE_k) that works though education and the direct effect (DE_k), which goes though other channels. These components can be further decomposed. In particular, it is useful to separate parts of the direct effect that are explained though developmental mediators: skills and health.

$$\begin{aligned} IE_k =& (\tilde{d}_{8k} + \tilde{d}_{9k}SES)(\sum_s \tilde{c}^s_{51}(b^s_1 + b^s_3 SES) \\ &+ (\tilde{c}_{61} + \tilde{c}_{71}SES)(a_1 + a_3 SES) + (\tilde{c}_{11} + \tilde{c}_{31}SES)) \\ &+ (\tilde{d}_{10k} + \tilde{d}_{11k}SES)(\sum_s \tilde{c}^s_{53}(b^s_1 + b^s_3 SES) \\ &+ (\tilde{c}_{63} + \tilde{c}_{73}SES)(a_1 + a_3 SES) + (\tilde{c}_{13} + \tilde{c}_{33}SES)) \end{aligned} \tag{7}$$

$$DE_k = \underbrace{\sum_s \tilde{d}^s_{5k}(b^s_1 + b^s_3 SES)}_{\text{early skills}} + \underbrace{(\tilde{d}_{6k} + \tilde{d}_{7k}SES)(a_1 + a_3 SES)}_{\text{early health}} + \underbrace{(\tilde{d}_{1k} + \tilde{d}_{3k}SES).}_{\text{other channels}} \tag{8}$$

Here tildes above coefficients denote marginal coefficients in logit models, which are proportional to corresponding model coefficients.

We impute missing control variables and factor scores to increase estimation sample size and statistical power using the well-established MCMC multiple imputation procedure, which is known to preserve the variance-covariance matrix of variables (Schafer, 1999).

5 Results

Below we first explore the interaction of the education PGS and SES using a reduced-form regression comparable with the literature. Then we analyse results of the main model, which controls for a full set of background controls, cognitive and noncognitive skills, early health, and unobserved heterogeneity. Results of the main model are consistent with results of the reduced-form regression and provide insights about the mechanisms behind the observed effects.

Exploring Interactions using a Reduced-Form Regression Here we estimate model (1) to explore the associations between health outcomes, the education PGS, and the interaction between PGS and SES. Our main SES measure is similar to the one used in Bierut et al. (2018) in that we summarize three qualitative measures of major socioeconomic adversities, with the difference being that we use a binary variable instead of an index taking values 0, 1, 2, and 3 to avoid model misspecification induced by a possible nonlinear effects of the SES index.

Table 3 presents results for model (1) using two alternative measures of SES. In Panel A we use the low SES dummy. In Panel B we use the logarithm of family income.

In Panel A of Table 3, the education PGS consistently shows statistically significant and strong associations with health-related outcomes. An exception to this pattern is the lack of a relationship between the education PGS and suicidal thoughts.

As expected, low SES is adversely related to health, while income is positively associated with health. This confirms a well-known relationship, and we show that this relationship survives conditioning on genetic endowment related to education, measured by the education PGS. Note that even though income has the advantage of being continuous and is not an ad hoc measure of SES, its predictive power is inferior to that of our binary SES measure. Income does not predict marijuana use or depression, while the binary variable does. This difference may reflect the fact that the low SES variable specif-

Table 3: Conditional Reduced-Form Associations Between Education PGS, Gene-Environment Interactions, and Health-Related Outcomes

	Health (1)	Heavy Drinking (2)	Smoking Tobacco (3)	Marijuana Use (4)	Obese (5)	No Exercise (6)	Depression (7)	Suicidal Thoughts (8)
A) Interaction with SES								
Education PGS	0.109 ***	-0.027 ***	-0.048 ***	-0.018 ***	-0.031 ***	-0.019 ***	-0.025 ***	0.004
	(0.019)	*(0.009)*	*(0.010)*	*(0.005)*	*(0.010)*	*(0.007)*	*(0.008)*	*(0.005)*
Poor SES	-0.278 ***	0.037 **	0.069 ***	0.028 ***	0.067 ***	0.030 **	0.027 **	0.029 ***
	(0.035)	*(0.015)*	*(0.018)*	*(0.008)*	*(0.017)*	*(0.012)*	*(0.013)*	*(0.008)*
Education PGS x Poor SES	-0.082 **	0.022	-0.005	0.018 **	0.003	0.023 *	0.026 *	-0.008
	(0.035)	*(0.015)*	*(0.018)*	*(0.008)*	*(0.017)*	*(0.012)*	*(0.013)*	*(0.008)*
Sample size	3813	3795	3813	3810	3768	3654	3813	3804
B) Interaction with Income								
Education PGS	0.074 ***	-0.022 ***	-0.051 ***	-0.012 ***	-0.020 **	-0.008	-0.019 ***	0.000
	(0.017)	*(0.008)*	*(0.009)*	*(0.004)*	*(0.008)*	*(0.006)*	*(0.007)*	*(0.004)*
Log of family income	0.121 ***	-0.017 **	-0.027 ***	-0.007	-0.041 ***	-0.012 **	0.003	-0.008 *
	(0.017)	*(0.007)*	*(0.009)*	*(0.004)*	*(0.008)*	*(0.006)*	*(0.007)*	*(0.004)*
Education PGS x log of family income	-0.007	-0.012	-0.002	-0.007 *	-0.001	-0.007	-0.009	-0.002
	(0.016)	*(0.007)*	*(0.009)*	*(0.004)*	*(0.008)*	*(0.006)*	*(0.007)*	*(0.004)*
Sample size	3454	3439	3454	3451	3414	3311	3454	3445

Notes: Calculations are based on the Add Health data. Reported coefficients for binary health outcomes (2–8) are estimated marginal effects of a logit model. All results are conditional on a limited set of controls comparable to those used in the literature: principal components 1–5 for ancestry, binary variables for US regions, cohorts, and sex. Asterisks indicate statistical significance level: ***, 1 % level; **, 5 % level; *, 10 % level.

ically captures severe economic hardship, while income captures many other margins of well-being, which are potentially less relevant.

Finally, in four out of eight cases, we can see a statistically significant and strong interaction between the education PGS and SES (see Panel A). Additionally, the interaction for heavy drinking is borderline statistically significant and the estimate of the coefficient has a sign consistent with the other interactions. The signs of these interaction effects are always such that interaction with low SES cancels the health-beneficial effect of the education PGS, a result to be confirmed by using our main model. Our results complement those by Bierut et al. (2018), as we find similar interaction effects but for a different type of PGS (we use the education PGS, not smoking PGS) and different outcomes (we use a variety of health-related outcomes, not only smoking tobacco). This study also complements the results of Papageorge and Thom (2018), who use an interaction with childhood SES to study education as an outcome.

In contrast with the low SES dummy, the interaction with income provides only one interaction effect, which is statistically significant at the 10% level. This result is consistent with the bottleneck hypothesis (Fletcher, 2019): low SES provides a good approximation for severely constrained conditions, whereas a substantial part of log income variation is outside the area of severe environmental constraints.

We also explore family fixed effect estimates documented in Table A-2 of the Appendix, but find out that they suffer from low statistical power, the same result that Amin et al. (2019) have based on the same dataset but a model using different PGS score and different outcome. Even the effect of the education PGS on education is not precisely determined. Statistical power is greatly diminished compared to our main model for two reasons: (1) restriction of identifying variation to within family variation and (2) a dramatic decrease in the sample size: only about 230 families who have at least two children surveyed by Add Health contribute to the estimation.

The Effect of the Education PGS on Health-Related Outcomes by SES All results discussed from this point are based on the main model (2–6), which allows for superior controls compared to the reduced-form regression model (1). The superior controls include a substantially larger set of observable background controls, the cognitive factor, multiple noncognitive factors, early health, and unobserved heterogeneity. Under the assumptions of the model the results can be viewed as causal. A conservative interpretation of the results would be as associations conditional on a large number of essential observables and unobservables.

We start with aggregated results. Table 4 shows the total effect of the education PGS on education and health-related outcomes by SES level and a decomposition of this effect with respect to education into direct and indirect effects based on formulas (7) and (8).

Columns (1) and (2) show beneficial effects of the education PGS on education, which are similar across SES level (see Panel C for no statistically significant differences in effects by SES level). For six health-related outcomes (health, heavy drinking, marijuana use, obesity, no exercise, and depression) we can see statistically significant total effects of the education PGS for high SES and no such effects for low SES (compare total effects in Panel A and Panel B). The corresponding regression coefficients tend to be large for high SES and small for low SES[11], which suggests that statistically insignificant effects for low SES are driven not by standard deviation of estimators but by either nonexistent or negligible effects. The difference in total effects by SES for these outcomes, shown in Panel C, tends to be either statistically significant or borderline statistically significant. These results confirm the substantial gene-SES interaction that favors the high-SES individuals, which we have seen based on the reduced form model.[12]

Further, for each SES level, we decompose the total effect into an indirect effect (work-

[11]Obesity is the only exception where the statistically insignificant estimate is large in magnitude.

[12]The only exception is smoking, for which we can see statistically significant and comparable effects for both levels of SES. Also, we see no effects for both SES levels for the "suicidal thoughts" variable.

Table 4: Direct, Indirect, and Total Effects of Education PGS on Health-Related Outcomes

	Bachelor's or above (1)	High school dropout (2)	Health (3)	Heavy Drinking (4)	Smoking Tobacco (5)	Marijuana Use (6)	Obese (7)	No Exercise (8)	Depression (9)	Suicidial Thoughts (10)
A. High SES										
Indirect effect	-	-	0.017 ***	-0.007 ***	-0.019 ***	-0.005 ***	-0.005 **	-0.004 ***	-0.005 ***	0.000
	-	-	*(0.004)*	*(0.002)*	*(0.003)*	*(0.001)*	*(0.002)*	*(0.001)*	*(0.002)*	*(0.001)*
Direct effect	-	-	0.064 ***	-0.011	-0.031 ***	-0.010 **	-0.014	-0.010	-0.020 ***	0.004
	-	-	*(0.019)*	*(0.009)*	*(0.011)*	*(0.004)*	*(0.010)*	*(0.007)*	*(0.008)*	*(0.004)*
Total	0.069 ***	-0.006 *	0.081 ***	-0.018 **	-0.050 ***	-0.015 ***	-0.019 *	-0.014 **	-0.026 ***	0.004
	(0.011)	*(0.003)*	*(0.019)*	*(0.009)*	*(0.011)*	*(0.004)*	*(0.010)*	*(0.007)*	*(0.008)*	*(0.004)*
B. Low SES										
Indirect effect	-	-	0.029 ***	-0.006 *	-0.019 ***	-0.005 **	-0.003	-0.006 **	0.001	-0.001
	-	-	*(0.009)*	*(0.003)*	*(0.005)*	*(0.002)*	*(0.004)*	*(0.003)*	*(0.003)*	*(0.002)*
Direct effect	-	-	-0.022	0.004	-0.047 ***	0.006	-0.019	0.010	-0.003	-0.004
	-	-	*(0.031)*	*(0.013)*	*(0.016)*	*(0.005)*	*(0.014)*	*(0.009)*	*(0.010)*	*(0.006)*
Total	0.074 ***	-0.012 **	0.006	-0.002	-0.066 ***	0.001	-0.022	0.003	-0.002	-0.005
	(0.018)	*(0.005)*	*(0.030)*	*(0.013)*	*(0.017)*	*(0.005)*	*(0.014)*	*(0.009)*	*(0.010)*	*(0.006)*
C. Difference[(a)]										
Δ Indirect	-	-	0.011	0.001	0.000	0.000	0.002	-0.002	0.006 **	-0.001
	-	-	*(0.010)*	*(0.004)*	*(0.006)*	*(0.002)*	*(0.004)*	*(0.003)*	*(0.003)*	*(0.002)*
Δ Direct	-	-	-0.086 **	0.016	-0.016	0.016 **	-0.005	0.020 *	0.018	-0.008
	-	-	*(0.036)*	*(0.016)*	*(0.019)*	*(0.007)*	*(0.018)*	*(0.011)*	*(0.013)*	*(0.008)*
Δ Total	0.005	-0.006	-0.075 **	0.017	-0.016	0.016 **	-0.003	0.018	0.024 *	-0.009
	(0.021)	*(0.006)*	*(0.035)*	*(0.016)*	*(0.020)*	*(0.007)*	*(0.017)*	*(0.011)*	*(0.013)*	*(0.007)*

Notes: Calculations are based on the AddHealth data. All results are conditional on the complete set of controls. Asterisks indicate statistical significance level: ***, 1 % level; **, 5 % level; *, 10 % level. [(a)]Difference is between low and high SES estimates.

ing through education) and the direct effect (working through channels other than education). The indirect effect tends to be similar for high and low-SES individuals (compare indirect effects in Panels A and B), but the direct effect shows differences for a number of outcomes.

Panel C of Table 4 estimates and tests the differences by SES, which are interaction effects. Δ Total in Panel C of Table 4 is an analogue of the interaction term in Panel A of Table 3. The statistically significant reduced-form interaction terms are 8-28% larger than the corresponding and more reliable results from Panel C of Table 4. This difference should be the result of controlling for a larger set of observables, for multiple skills, and for unobserved heterogeneity in the main model on which Table 4 is based. The level of statistical significance is preserved relative to Table 3 for all outcomes except for "no exercise," for which a statistically significant interaction effect at the 10% level becomes borderline statistically significant based on the main model.

We can see from Panel C of Table 4 that the only statistically significant interaction for the indirect effect can be observed for depression, which means that education induced by genetic endowments tends to contribute to health in a similar way by SES.

In contrast, for the direct effect the role of the interaction tends to be important. The interaction is statistically significant for general health, marijuana use, and no exercise (see Panel C). In addition, for depression, the interaction is borderline statistically significant, which corresponds to large and statistically significant direct effect for high SES (Panel A) and statistically insignificant effect for low SES individuals (Panel B).

To summarize, although we can see that interaction effects are smaller and less precisely determined than we would expect based on reduced-form analysis, the results are largely preserved when using the main model. Additionally, we learn that indirect effects, which are explained by education, tend to be similar across SES, while the direct effects, which are not explained by education, tend to be explained by the interaction between genetic endowments and SES. Perhaps, apart from education benefits,

having a high education PGS leads children and young adults to better use family resources for their development (parental knowledge and networks, family wealth, access to health care, access to certain compulsory schools), but high-SES children have superior resources to leverage their genetic endowment.

Decomposing Direct Effects Because skills and health are partly genetic (Borghans et al., 2008), we use formula (8) and decompose the direct effect of Education PGS, estimates of which we have seen in Table 4, with respect to early health, early skills, and other channels.

Table 5 shows that skills and health play some role, but a large part of the effect for high-SES individuals is explained by other channels. The other channels are not statistically significant for the low-SES group with the only exception of smoking.

We have two conclusions from these results: (1) The education PGS partly influences adult health through early health and skills; (2) There is a large part of the direct effect that is explained by SES and not by health or skills. This channel might be partly due to access to family resources discussed above and, partly, possible incomplete control for health and skills because of data limitations.

There are several unobserved skills that might add to the estimated contribution of "other channels." First, the education PGS may affect not only early health and early skills, but also risk conceptions in early life, which may not be fully captured by available measures of skills. Recent evidence suggests that a large portion of an individual's risk tolerance is heritable (Linner et al., 2019), and so it might be correlated with the education PGS. Second, it is possible that the education PGS may not only affect risk tolerance itself, but may also affect how individuals evaluate risk in the first place. Barth et al. (2018) show that an education PGS predicts whether individuals distinguish that scenarios have various degrees of probability. Individuals with higher polygenic scores may better understand risks associated with health behaviors. Third, the direct

Table 5: Decomposition of the Direct Effect of the Education PGS on Health-Related Outcomes

	Health (1)	Heavy Drinking (2)	Smoking Tobacco (3)	Marijuana Use (4)	Obese (5)	No Exercise (6)	Depression (7)	Suicidial Thoughts (8)
A. High SES								
Early health	0.006	0.000	0.000	0.000	-0.002	0.000	-0.001	0.000
	(0.004)	*(0.000)*	*(0.000)*	*(0.000)*	*(0.002)*	*(0.000)*	*(0.001)*	*(0.000)*
Skills	0.006 *	-0.003 *	-0.004 **	0.000	-0.002	-0.001	0.000	-0.001
	(0.004)	*(0.001)*	*(0.002)*	*(0.001)*	*(0.001)*	*(0.001)*	*(0.002)*	*(0.001)*
Other channels	0.052 ***	-0.009	-0.026 **	-0.010 **	-0.010	-0.009	-0.019 **	0.006
	(0.018)	*(0.009)*	*(0.011)*	*(0.004)*	*(0.010)*	*(0.007)*	*(0.008)*	*(0.004)*
Total direct effect	0.064 ***	-0.011	-0.031 ***	-0.010 **	-0.014	-0.010	-0.020 ***	0.004
	(0.019)	*(0.009)*	*(0.011)*	*(0.004)*	*(0.010)*	*(0.007)*	*(0.008)*	*(0.004)*
B. Low SES								
Early health	0.010 **	-0.001	-0.003 *	0.000	-0.003 *	-0.001	-0.002 *	0.000
	(0.005)	*(0.001)*	*(0.002)*	*(0.000)*	*(0.002)*	*(0.001)*	*(0.001)*	*(0.000)*
Skills	0.006	-0.002	-0.003	0.000	-0.002	-0.001	0.000	-0.001
	(0.004)	*(0.002)*	*(0.002)*	*(0.001)*	*(0.001)*	*(0.001)*	*(0.002)*	*(0.001)*
Other channels	-0.038	0.007	-0.040 **	0.006	-0.015	0.011	0.000	-0.002
	(0.030)	*(0.013)*	*(0.016)*	*(0.005)*	*(0.015)*	*(0.009)*	*(0.010)*	*(0.006)*
Total direct effect	-0.022	0.004	-0.047 ***	0.006	-0.019	0.010	-0.003	-0.004
	(0.031)	*(0.013)*	*(0.016)*	*(0.005)*	*(0.014)*	*(0.009)*	*(0.010)*	*(0.006)*

Notes: Calculations are based on the AddHealth data. All results are conditional on the complete set of controls. Asterisks indicate statistical significance level: ***, 1 % level; **, 5 % level; *, 10 % level.

effect might be partly explained by educational aspirations. Belsky et al. (2016) find that the education PGS is associated with higher educational aspirations, even from a young age. It may be possible that the education PGS captured elements such as self-esteem, confidence, or social ambition, which contribute to both superior health and better education. Finally, early Agreeableness and Openness were not available in the data set and can explain a part of the effect though other channels.

Evidence for the Causal Effect of Education on Health and Education-SES Interaction

Table 6 shows the marginal conditional associations between high school dropout status, completion of a bachelor's degree or above, and each health outcome. Associations are relative to high school education, an omitted binary variable. In Panel (A) we find that Bachelor's degree is associated with better general health and a smaller likelihood of adverse health behaviors. Only "having suicidal thoughts" among the studied outcomes is not associated with Bachelor's degree. In contrast, high school dropouts have worse general health than high school graduates. Additionally, they are more likely to smoke tobacco, use marijuana, and be depressed.

Another result of Table 6 is the interaction between education levels and low SES status (see Panel (C)). While statistically significant interaction is observed only for a few variables, it is notable that general health is among them. General health is important, as it summarizes multiple inputs to the health production function including those which are unobserved. Results suggest that Bachelor's degree has higher health returns to low-SES individuals while dropping out of high school implies a sizable drop in health for high-SES, but not for low-SES individuals.

One possible interpretation of this result is that the average health and lifestyles of high-SES high school graduates are already sufficiently high due to parental investments and the high quality of their compulsory schooling, and so graduation from college contributes less to their health than it contributes to health of low-SES high school graduates,

Table 6: Marginal Effects of Education and Health-Related Outcomes in Young Adulthood

	Health (1)	Heavy Drinking (2)	Smoking Tobacco (3)	Marijuana Use (4)	Obese (5)	No Exercise (6)	Depression (7)	Suicidial Thoughts (8)
A. High SES								
Bachelor's Degree or above[a]	0.175 ***	-0.083 ***	-0.198 ***	-0.057 ***	-0.064 ***	-0.053 ***	-0.056 ***	0.006
	(0.042)	*(0.021)*	*(0.025)*	*(0.011)*	*(0.024)*	*(0.017)*	*(0.019)*	*(0.011)*
High School Dropout	-0.278 ***	0.027	0.282 ***	0.024 *	-0.014	-0.001	0.069 **	0.021
	(0.077)	*(0.033)*	*(0.044)*	*(0.013)*	*(0.038)*	*(0.019)*	*(0.027)*	*(0.016)*
B. Low SES								
Bachelor's Degree or above	0.329 ***	-0.070 *	-0.190 ***	-0.047 **	-0.055	-0.076 **	0.013	-0.015
	(0.079)	*(0.037)*	*(0.045)*	*(0.021)*	*(0.040)*	*(0.031)*	*(0.030)*	*(0.019)*
High School Dropout	-0.003	-0.012	0.194 ***	0.032 **	-0.110 ***	-0.029	0.040	-0.021
	(0.076)	*(0.032)*	*(0.041)*	*(0.013)*	*(0.036)*	*(0.024)*	*(0.026)*	*(0.015)*
C. Difference[b]								
Δ Bachelor's Degree or above	0.154 *	0.013	0.008	0.010	0.009	-0.023	0.069 **	-0.021
	(0.085)	*(0.041)*	*(0.050)*	*(0.023)*	*(0.045)*	*(0.034)*	*(0.034)*	*(0.022)*
Δ High School Dropout	0.275 ***	-0.040	-0.088	0.008	-0.096 *	-0.028	-0.030	-0.041 *
	(0.105)	*(0.044)*	*(0.059)*	*(0.017)*	*(0.051)*	*(0.029)*	*(0.036)*	*(0.022)*

Notes: Calculations are based on the AddHealth data. All results are conditional on the complete set of controls of the main model including early skills and health, Education PGS, and unobserved heterogeneity. Asterisks indicate statistical significance level: ***, 1 % level; **, 5 % level; *, 10 % level. [a]"High school graduate" is the omitted binary variable. [b]Difference is between low and high SES estimates.

for whom college graduation opens new opportunities in terms of lifestyles and health care quality. In contrast, dropping out of high school puts high-SES individuals into a position with very limited options for success in life and adversely affects their health and lifestyle relative to high-SES high school graduates. In contrast, for low-SES individuals, dropping out of high school might have smaller consequences health-wise, as schools that they attend are generally low quality anyways (due to segregation caused by housing prices).

Another result of Table 6 is that completing a Bachelor's degree reduces depression for high-SES students but not for low-SES students. Perhaps socioeconomic advancement is associated with additional stress. Also, dropping out of high school increases depression for high-SES students but not for low-SES students. Socioeconomic decline might be stressful too.

The main contribution of this table though is that associations between education and health survive conditioning for a major genetic confounder: the education PGS. Historically, genes were part of unobservables in studies of education-health relationships, and controlling for a variable that summarizes hundreds of relevant genes was unimaginable until recently. These results contribute to resolving the controversy about the causal effect of education on health.

6 Conclusions

We find that an education polygenic score exhibits economically relevant effects on a variety of health outcomes both indirectly through education and directly. However, these effects are substantially moderated by SES, such that individuals who grew up in low-SES households do not experience the full health benefits of the education PGS. Genes are unlikely to be a policy variable in the foreseeable future due to ethical, political, and practical considerations, but we provide evidence that targeting the SES bottleneck may

allow individuals to achieve better health based on their genetic endowment. Based on the evidence here, targeting individuals at the low end of the SES spectrum appears to be an efficient way of boosting health outcomes. We contribute to an understanding of the mechanisms through which the education PGS works and provide a mediation model that can be used to guide future research. We also provide additional evidence that college completion has a causal effect on health.

References

Aitken, M. and D. B. Rubin (1985). Estimation and hypothesis testing in finite mixture models. *Journal of Royal Statistical Society, Series B 47 (1)*, pp. 67–75.

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

Amin, V., J. Behrman, J. M. Fletcher, C. A. Flores, A. Flores-Lagunes, and H.-P. Kohler (2019). Mental health, schooling attainment and polygenic scores: Are there significant gene-environment associations?

Amin, V., J. R. Behrman, and H.-P. Kohler (2015). Schooling has smaller or insignificant effects on adult health in the US than suggested by cross-sectional associations: New estimates using relatively large samples of identical twins. *Social Science & Medicine 127*, 181–189.

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.

Ashenfelter, O. and A. Krueger (1994). Estimates of the economic return to schooling from a new sample of twins. *American Economic Review 84*, 1157–1173.

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

Avinun, R. (2019). Educational attainment polygenic score is associated with depressive symptoms via socioeconomic status: A gene-environment-trait correlation. *bioRxiv*, 727552.

Barcellos, S. H., L. S. Carvalho, and P. Turley (2018). Education can reduce health differences related to genetic risk of obesity. *Proceedings of the National Academy of Sciences 115*(42), E9765–E9772.

Barth, D., N. W. Papageorge, and K. Thom (2018). Genetic endowments and wealth inequality. Technical report, National Bureau of Economic Research. Accepted at *Journal of Political Economy*.

Behrman, J. R., H.-P. Kohler, V. M. Jensen, D. Pedersen, I. Petersen, P. Bingley, and K. Christensen (2011). Does more schooling reduce hospitalization and delay mortality? New evidence based on danish twins. *Demography* (48), 1347–1375.

Belsky, D. W., T. E. Moffitt, D. L. Corcoran, B. Domingue, H. Harrington, S. Hogan, R. Houts, S. Ramrakha, K. Sugden, B. S. Williams, et al. (2016). The genetics of success: How single-nucleotide polymorphisms associated with educational attainment relate to life-course development. *Psychological science 27*(7), 957–972.

Benjamin, D. J., D. Cesarini, C. F. Chabris, E. L. Glaeser, D. I. Laibson, G. Age, V. Gudhnason, T. B. Harris, L. J. Launer, S. Purcell, et al. (2011). The promises and pitfalls of genoeconomics. *Annual Reviews of Economics 4*(1), 627–662.

Bierut, L., P. Biroli, T. Galama, and K. Thom (2018). Childhood socioeconomic status moderates genetic predisposition for peak smoking. *bioRxiv*. doi: http://dx.doi.org/10.1101/336834.

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

Boardman, J. D., B. W. Domingue, and J. Daw (2015). What can genes tell us about the relationship between education and health? *Social Science & Medicine 127*, 171–180.

Böckerman, P., J. Viinikainen, L. Pulkki-Råback, C. Hakulinen, N. Pitkänen, T. Lehtimäki, J. Pehkonen, and O. T. Raitakari (2017). Does higher education protect against obesity? evidence using mendelian randomization. *Preventive medicine 101*, 195–198.

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

Cameron, A. C. and P. K. Trivedi (2005). *Microeconometrics: Methods and applications*. Cambridge university press.

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.

Conti, G. and J. J. Heckman (2010). Understanding the early origins of the education-health gradient: A framework that can also be applied to analyze gene-environment interactions. *Perspectives on Psychological Science 5*(5), 585–605.

Conti, G., J. J. Heckman, and R. Pinto (2016). The effects of two influential early childhood interventions on health and healthy behaviour. *Economic Journal 126*, 28–65.

Conti, G., J. J. Heckman, and S. Urzúa (2010). The education-health gradient. *American Economic Review: Papers and Proceedings 100*(2), 234–238.

Cunha, F. and J. Heckman (2007). The technology of skill formation. *American Economic Review 97*(2), 31–47.

Davies, N. M., M. Dickson, G. Davey Smith, F. Windmeijer, and G. J. van den Berg (2019). The causal effects of education on adult health, mortality and income: Evidence from mendelian randomization and the raising of the school leaving age.

DiPrete, T. A., C. A. Burik, and P. D. Koellinger (2018). Genetic instrumental variable regression: Explaining socioeconomic and health outcomes in nonexperimental data. *Proceedings of the National Academy of Sciences 115*(22), E4970–E4979.

Domingue, B. W., D. W. Belsky, D. Conley, K. M. Harris, and J. D. Boardman (2015). Polygenic influence on educational attainment: New evidence from the National Longitudinal Study of Adolescent to Adult Health. *AERA Open 1*(3), 1–13.

Fletcher, J. M. (2019). Environmental bottlenecks in children's genetic potential for adult socio-economic attainments: Evidence from a health shock. *Population Studies 73*(1), 139–148.

Fletcher, J. M. and S. Schurer (2017). Origins of adulthood personality: The role of adverse childhood experiences. *The BE Journal of Economic Analysis & Policy 17*(2).

Freudenberg, N. and J. Ruglis (2007). Reframing school dropout as a public health issue. *Preventing chronic disease 4*(4).

Galama, T., A. Lleras-Muney, and H. van Kippersluis (2018). The effect of education on health and mortality: A review of experimental and quasi-experimental evidence. *Oxford Research Encyclopedias: Economics and Finance*.

Galama, T. J. and H. Van Kippersluis (2018). A theory of socio-economic disparities in health over the life cycle. *The Economic Journal 129*(617), 338–374.

Gilleskie, D. (2014). *Encyclopedia of Health Economics, Volume 1*, Chapter "Dynamic models: econometric considerations of time", pp. 209–216. Elsevier Inc.

Grossman, M. (2006). *Handbook of the Economics of Education*, Chapter "Education and nonmarket outcomes", pp. 577–633. Elsevier.

Harris, K. M. (2013). The Add Health study: Design and accomplishments. *Chapel Hill: Carolina Population Center, University of North Carolina at Chapel Hill*.

Harris, K. M. and D. B. Braudt (2018). Polygenic scores (PGSs) in the National Longitudinal Study of Adolescent to Adult Health (Add Health)–release.

Heckman, J. and B. Singer (1984). A method for minimizing the impact of distributional assumptions in econometric models for duration data. *Econometrica: Journal of the Econometric Society*, 271–320.

Heckman, J. J. (2007). The economics, technology, and neuroscience of human capability formation. *Proceedings of the national Academy of Sciences 104*(33), 13250–13255.

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 (2019). Understanding the mechanisms linking education with longevity. SSRN Working Paper http://dx.doi.org/10.2139/ssrn.2638735.

Idler, E. L. and Y. Benyamini (1997). Self-rated health and mortality: a review of twenty-seven community studies. *Journal of Health and Social Behavior*, 21–37.

Lee, J. J., R. Wedow, A. Okbay, E. Kong, O. Maghzian, M. Zacher, T. A. Nguyen-Viet, P. Bowers, J. Sidorenko, R. K. Linnér, et al. (2018). Gene discovery and polygenic prediction from a genome-wide association study of educational attainment in 1.1 million individuals. *Nature Genetics 50*(8), 1112–1121.

Linner, R. K., P. Biroli, E. Kong, S. F. W. Meddens, R. Wedow, M. A. Fontana, and et al. (2019). Genome-wide association analyses of risk tolerance and risky behaviors in over 1 million individuals identify hundreds of loci and shared genetic influences. *Nature Genetics 51*(2), 245–252.

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

Lundborg, P., C. H. Lyttkens, and P. Nystedt (2016). The effect of schooling on mortality: New evidence from 50,000 swedish twins. *Demography* (53), 1135–1168.

Maddala, G. (1983). *Limited-Dependent and Qualitative Variables in Econometrics*. New York: Cambridge University Press.

Madsen, M., A.-M. N. Andersen, K. Christensen, P. K. Andersen, and M. Osler (2010). Does educational status impact adult mortality in Denmark? A twin approach. *American Journal of Epidemiology 172*.

Marioni, R. E., S. J. Ritchie, P. K. Joshi, S. P. Hagenaars, A. Okbay, K. Fischer, M. J. Adams, W. D. Hill, G. Davies, R. Nagy, et al. (2016). Genetic variants linked to education predict longevity. *Proceedings of the National Academy of Sciences 113*(47), 13366–13371.

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

Meghir, C., M. Palme, and E. Simeonova (2018, April). Education and mortality: Evidence from a social experiment. *American Economic Journal: Applied Economics 10*(2), 234–56.

Mroz, T. A. (1999). Discrete factor approximations in simultaneous equation models: Estimating the impact of a dummy endogenous variable on a continuous outcome. *Journal of Econometrics 92*(2), 233–274.

Munafò, M., N. M. Davies, and G. D. Smith (2019). Can genetics reveal the causes and consequences of educational attainment? *Journal of the Royal Statistical Society: Series A (Statistics in Society)*.

National Institute on Alcohol Abuse and Alcoholism (2004). NIAAA council approves definition of binge drinking. Technical Report 3.

Papageorge, N. W. and K. Thom (2018). Genes, education, and labor market outcomes: evidence from the health and retirement study. Technical report, National Bureau of Economic Research. Accepted subject to minor revisions at *Journal of the European Economic Association*.

Price, A. L., N. J. Patterson, R. M. Plenge, M. E. Weinblatt, N. A. Shadick, and D. Reich (2006). Principal components analysis corrects for stratification in genome-wide association studies. *Nature genetics 38*(8), 904–909.

Savelyev, P., B. Ward, R. Krueger, and M. McGue (2019). Health endowments, schooling allocation in the family, and longevity: Evidence from US twins. SSRN Working Paper http://dx.doi.org/10.2139/ssrn.3193396.

Savelyev, P. A. (2019). Conscientiousness, Extraversion, college education, and longevity of high-ability individuals. SSRN Working Paper http://dx.doi.org/10.2139/ssrn.1715942.

Savelyev, P. A. and K. T. Tan (2019). Socioemotional skills, education, and health-related outcomes of high-ability individuals. *American Journal of Health Economics 5*, pp. 250–280.

Schafer, J. L. (1999). Multiple imputation: a primer. *Statistical methods in medical research 8*(1), 3–15.

Schmitz, L. L. and D. Conley (2017). The effect of vietnam-era conscription and genetic potential for educational attainment on schooling outcomes. *Economics of Education Review 61*, 85–97.

Strauss, J. and D. Thomas (2007). Health over the life course. *Handbook of Development Economics 4*, 3375–3474.

Tillmann, T., J. Vaucher, A. Okbay, H. Pikhart, A. Peasey, R. Kubinova, A. Pajak, A. Tamosiunas, S. Malyutina, F. P. Hartwig, et al. (2017). Education and coronary heart disease: mendelian randomisation study. *bmj 358*, j3542.

van den Berg, G., L. Janys, and K. Christensen (2015). The effect of education on mortality. IZA working paper.

van den Berg, G. J. (2001). Duration models: Specification, identification and multiple durations. In J. J. Heckman and E. E. Leamer (Eds.), *Handbook of Econometrics*, Volume 5 of *Handbooks in Economics*, pp. 3381–3460. New York: North-Holland.

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.

von Hinke, S., G. D. Smith, D. A. Lawlor, C. Propper, and F. Windmeijer (2016). Genetic markers as instrumental variables. *Journal of Health Economics 45*, 131–148.

Young, J. K. and A. A. Beaujean (2011). Measuring personality in wave I of the national longitudinal study of adolescent health. *Frontiers in Psychology 2*.

Yung, Y.-F. and K.-H. Yuan (2013). Bartlett factor scores: General formulas and applications to structural equation models. In R. E. Millsap, L. A. van der Ark, D. M. Bolt, and C. M. Woods (Eds.), *New Developments in Quantitative Psychology*, New York, NY, pp. 385–401. Springer New York.

A Appendix

Table A-1: Early Cognitive and Noncognitive Measures

Conscientiousness	Emotional Stability
Gathers facts when solving problems	Has good qualities
Thinks of alternative ways to solve problems	Has a lot to be proud of
Uses systematic methods when solving problems	Likes oneself
Analyzes outcome of solutions to problems	Feels like doing things right
	Feels socially accepted
	Feels loved and wanted
Extraversion	Cognition
Feels close to people at school	Add Health Picture Vocabulary Test
Feels like a part of the school	Recent math grade
Feels socially accepted	Recent science grade

Note: Measures are self-reported and taken from the first wave of the Add Health data. The set of early noncognitive measures is taken from the analysis by Young and Beaujean (2011). Scores for math and science are imputed from letter grades.

Table A-2: Conditional Reduced-Form Associations Between Education PGS, Gene-Environment Interactions, and Health-Related Outcomes, an Alternative Fixed Effects Estimation, which Suffers from Low Statistical Power

	Bachelor (1)	Dropout (2)	Health (3)	Heavy Drinking (4)	Smoking Tobacco (5)	Marijuana Use (6)	Obese (7)	No Exercise (8)	Depression (9)	Suicidal Thoughts (10)
Education PGS	0.131	-0.006	0.004	0.054	-0.037	-0.006	0.012	-0.001	-0.030	0.012
	(0.084)	*(0.067)*	*(0.310)*	*(0.143)*	*(0.158)*	*(0.060)*	*(0.174)*	*(0.160)*	*(0.146)*	*(0.103)*
Education PGS x low SES	0.015	-0.007	-0.126	-0.075	-0.078	0.014	-0.101	-0.074	-0.027	-0.050
	(0.123)	*(0.125)*	*(0.369)*	*(0.185)*	*(0.161)*	*(0.096)*	*(0.184)*	*(0.180)*	*(0.147)*	*(0.128)*
Number of individuals	463	463	463	461	463	462	458	437	463	461

Notes: Calculations are based on 230 families with at least two children from the AddHealth data. Lack of asterisks in the table corresponds to no statistically significant effects at the 10% level. We exclude identical twins, as they share the same PGS and SES. All regressions are conditional on the following regressors that may differ across children from the same family: 1^{st}, 2^{d}, and 3^{d}-born, meals with parents, low birth weight, five genetics ancestry principle components, age, and sex.